TARGETS

SEE FOLLOWING NOTE

WE REGRET THAT THIS DOCUMENT MAY BE INCOMPLETE AND/OR PARTIALLY ILLEGIBLE.

SINCE THERE ARE NO BETTER COPIES AVAILABLE, MICROMEDIA HAS DECIDED TO FILM THE DOCUMENT FOR ARCHIVAL PURPOSES.

VOIR NOTE CI-APRES

NOUS REGRETTONS QUE CE DOCUMENT N'EST PAS COMPLETE ET/OU EST EN PARTIE ILLISIBLE.

MEILLEURS COPIES N'EXISTENT PAS DONC MICROMEDIA A DECIDE DE FILMER LE DOCUMENT POUR LES BESOINS DES ARCHIVES.

micromedia
a division of IHS Canada

20 Victoria Street
Toronto, Ontario M5C 2N8
Tel.: (416) 362-5211
Toll Free: 1-800-387-2689
Fax: (416) 362-6161
Email: info@micromedia.on.ca

CRT-2000-36

Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

THE CELL-TRANSMISSSION MODEL: A NEW LOOK AT A DYNAMIC NETWORK LOADING MODEL

by

Shane M. Velan

Thèse de doctorat

November 2000

101-01019/

PUBLICATION
CRT-2000-36

THE CELL-TRANSMISSSION MODEL: A NEW LOOK AT A DYNAMIC NETWORK LOADING MODEL*

by

Shane M. Velan[1]

* Thèse présentée en vue de l'obtention du diplôme de Ph.D. Sous la direction de Michael Florian. Ce mémoire a été rendu possible grâce en partie au Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG).

[1] Département d'informatique et de recherche opérationnelle et Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7

Département d'informatique et de recherche opérationnelle - Publication #1185

©Shane M. Velan, 2000

Dépôt légal – Bibliothèque nationale du Québec
Bibliothèque nationale du Canada, 2000

Sommaire

(An English abstract follows this section.)

Étant donné un réseau de trafic composé d'arcs et de nœuds ainsi que la nature temporelle de la demande origine-destination, le problème de chargement dynamique des réseaux s'intéresse au flot temporel, à la densité et au temps de déplacement. Le modèle de transmission de cellule est une solution à résolution moyenne au problème de chargement dynamique des réseaux puisqu'il est plus perfectionné que les modèles analytiques, mais il n'est pas aussi détaillé que les modèles microscopiques. L'objectif premier est de mesurer, tant quantitativement que qualitativement, la précision avec laquelle le modèle de transmission de cellule résout le problème de chargement dynamique des réseaux.

La première étape a été de définir et de développer le modèle de transmission de cellule, ce qui n'a été possible qu'en ayant tout appris sur les bases théoriques du modèle. L'historique du modèle a été présenté, les contributions des différents domaines ont été intégrées, les lacunes dans la documentation ont été comblées et quelques réflexions ont été apportées.

Le modèle de transmission de cellule a été implémenté dans un programme de simulation appelé CellNetLoad. Le programme a servi à évaluer, sur des réseaux d'essai, la fidélité du modèle de transmission de cellule par comparaison avec un programme de simulation microscopique. En raison des résultats obtenus avec CellNetLoad, des améliorations ont été apportées au modèle. D'autres tests ont ensuite été effectués en utilisant les mêmes réseaux. Voici quelques-unes des améliorations apportées à la propagation du flot : relation flot-densité non linéaire, compensation de l'erreur dans la capacité d'entreposage des arcs et discrétisation du flot. Parmi les améliorations au modèle de carrefour, il y a ce qui suit : l'assouplissement de la discipline du premier entré premier sorti et la modélisation explicite du marquage des voies à l'approche des carrefours. Enfin, le programme CellNetLoad a été utilisé pour résoudre le problème de chargement

dynamique des réseaux dans un corridor de déplacement sur une autoroute unidirectionnelle reliée à une artère parallèle par le moyen d'une bretelle d'accès et d'une bretelle de sortie.

Il a été constaté que le modèle de transmission de cellule pouvait fidèlement propager, sur un arc, le flot, les vagues cinématiques implicites, les ondes de choc et les éventails et ce, s'il y a peu de tricotage (*weaving*). Il est possible d'assouplir l'hypothèse d'une relation trapézoïdale entre le flot et la densité, mais la fonction doit demeurer concave. La discrétisation du flot a causé un biais qui a limité le flot et elle n'est donc pas recommandée.

Le flot sur l'arc dépend des conditions limites aux nœuds. Si le modèle du nœud ne produit pas un flot exact aux carrefours, le flot des arcs ne sera pas exact. Le modèle de transmission de cellule respecte la discipline du premier entré premier sorti, ce qui n'est pas réaliste aux abords de carrefours à voies multiples. Les améliorations apportées aux algorithmes des carrefours ont rendu plus exactes les simulations des abords de carrefours à voies multiples.

Malheureusement, le modèle de transmission de cellule n'a pas la structure nécessaire pour modeliser explicitement la limitation de la capacité et les retards causés par les changements de voie et les mouvements opposés permis aux carrefours. De tels effets peuvent être prédits et intégrés à la mise à exécution du modèle. Cependant, il peut s'avérer difficile de trouver des prédictions justes. Par conséquent, il y a des limites au réalisme du modèle de transmission de cellule.

Abstract

Given a traffic network of arcs and nodes, and the temporal origin-destination demand, the dynamic network loading (DNL) problem seeks the temporal flow, density, and travel time. The cell-transmission model falls within a category of medium-resolution solutions to the DNL problem, since it is more refined than analytical models, but not as detailed as microscopic models. The primary objective is to determine quantitatively and qualitatively how accurately the cell-transmission model solves the DNL problem.

The first task is to define and develop the cell-transmission model, which requires a profound understanding of the theoretical foundations of the model. The necessary background was summarised, the contributions from a variety of fields were integrated, gaps in the literature were bridged, and insights were offered.

Next, the cell-transmission model was implemented in a computer simulation program, called CellNetLoad. This program was used to evaluate the accuracy of the cell-transmission model relative to a microscopic simulation program on a set of test networks. Based on the performance of CellNetLoad, several enhancements were implemented, then evaluated on the same test networks. The enhancements to flow propagation included a non-linear equilibrium flow-density relationship, compensation for the error in the arc storage capacity, and the discretisation of flow. The enhancements to the intersection model involve the relaxation of the constraint that permits only diverge and merge intersections, the relaxation of first-in-first-out (FIFO) discipline, and the explicit modelling of lane striping at intersection approaches. Finally CellNetLoad was used to solve the DNL problem for a commuter corridor with a unidirectional freeway linked to a parallel arterial via an on-ramp and an off-ramp.

It was found that the cell-transmission model can accurately propagate flow, as well as the implicit kinematic waves, shockwaves, and fans, along an arc if there is little weaving. The assumption of a trapezoidal flow-density relationship may be relaxed, but the function must remain concave. The discretisation of flow resulted in a bias which restrained the flow, and so it is not recommended.

The flow on the arc depends on the boundary conditions at the nodes. If the node model does not produce accurate intersection flows, then the accuracy of the flows on the arcs will be compromised. The cell-transmission model respects FIFO discipline, which is unrealistic at intersections with multi-lane approaches. The enhanced intersection algorithms improved the accuracy of the simulation of multi-lane approaches.

Unfortunately, the cell-transmission model lacks the necessary structure to incorporate capacity reductions and delay due to lane changes and opposed permissive movements at an intersection. These effects may be predicted and incorporated in the model input, but finding accurate predictions is an intricate challenge. Therefore, the potential accuracy of the cell-transmission model is limited.

Keywords: dynamic traffic flow macroscopic model

Résumé

(An English executive summary follows this section.)

L'objectif premier de la présente recherche est d'offrir une explication détaillée de la théorie qui sous-tend le modèle de transmission de cellule comme solution au problème de chargement dynamique des réseaux. Le modèle de transmission de cellule a été élaboré à partir d'éléments provenant de nombreux domaines. Par conséquent, la documentation sur ce sujet est incomplète. De plus, des modifications ont été apportées au modèle qui contredisent des postulats fondamentaux de la théorie.

Le deuxième objectif de la présente thèse est d'évaluer le réalisme du modèle de transmission de cellule et d'examiner les améliorations qui pourraient le rendre encore plus précis. Par ailleurs, l'effet des améliorations sur la simulation de la vitesse d'exécution sera étudié, même si ce n'est, dans la présente recherche, que d'importance secondaire.

Un examen exhaustif de la documentation portant sur le modèle de transmission de cellule a été effectué. Le modèle est très récent, ayant d'abord été élaboré par C.F. Daganzo en 1994 pour tenter de résoudre le problème de chargement dynamique sur un réseau de trafic. Le modèle de transmission de cellule réunit les constatations préliminaires de nombreux domaines de recherche, dont le génie du transport (Greenshields, 1935), la théorie de l'écoulement de la circulation (Lighthill et Whitham, 1955 ; Richards, 1956), les méthodes numériques (Lax, 1954), et la dynamique des fluides (Godunov, 1959). Sans oublier le travail d'Ansorge (1990) et de Lebacque (1996).

La première étape sera de décrire le problème de chargement dynamique des réseaux et de présenter les solutions proposées. Le modèle de transmission de cellule repose sur quelques éléments-clés, dont les variables d'état, c'est-à-dire le flot, la densité et la vitesse. Ces derniers

seront définis de même que les relations qui les unissent. Une attention particulière sera cependant portée à la relation flot-densité. Par ailleurs, des exemples de relations flot-densité non linéaires, à régimes multiples et linéaires par morceaux seront fournis. Enfin, la loi de la conservation du flot sera élaborée.

Le modèle hydrodynamique LWR (Lighthill et Whitham, 1955 ; Richards, 1956) est défini à l'aide de la loi de la conversation du flot et de la relation flot-densité en équilibre et ce, en tant que modèle de premier ordre dérivé de la dynamique des fluides. Le modèle LWR est une méthode de solution macroscopique du problème dynamique de chargement des réseaux. Un problème de valeur initiale, où la densité initiale d'une voie routière à relation flot-densité en équilibre est connue, peut être résolu en utilisant le modèle peu robuste LWR. Toutefois, la solution n'est pas unique. Il faut donc introduire la condition d'entropie afin de pouvoir choisir la solution physique la plus appropriée. Il est possible de trouver la solution exacte au problème de valeur initiale par une méthode analytique appelée méthode des caractéristiques. En connaissant la densité et le flot, il est possible d'utiliser la relation fondamentale pour trouver la vitesse et, par conséquent, le temps de déplacement.

Les approximations par différences finies au modèle LWR sont étudiées dans le but de concevoir une solution plus facilement mécanisée au problème dynamique de changement des réseaux. Lax a été le premier, en 1954, à proposer une méthode rudimentaire de différenciation. En 1959, Godunov a conçu une solution approximative. Celle-ci a été adaptée aux réseaux de trafic par Lebacque en 1996. Le problème de valeur initiale est ramené à une série de problèmes généralisés de Riemann. Lebacque (1996) a fait l'analogie entre la méthode Godunov et les concepts de demande locale et d'offre locale.

Daganzo (1994) a trouvé une solution simplifiée au problème généralisé de Riemann et ce, sous forme de demande locale minimum et d'offre locale minimum. Une fonction trapézoïdale est utilisée pour la relation flot-densité en équilibre. Le temps est discrétisé en une série d'intervalles de temps de même durée. Chaque arc est représenté par une série de cellules. Dans chaque cellule, la longueur est définie comme étant la distance parcourue durant un intervalle de temps par un véhicule qui se déplace à la vitesse à écoulement libre de l'arc. Il est démontré que la solution du modèle de transmission de cellule converge vers la solution du modèle LWR pour des discrétisations espace-temps plus précises.

Le modèle de transmission de cellule propose également un modèle de nœud. Cependant, seuls les carrefours de convergence et de bifurcation peuvent être considérés dans une logique de maximisation du flot. En raison de cette restriction, le flot de chaque cellule dépend uniquement de la densité de la cellule courante (demande locale) et de la densité des cellules (ou de la cellule) en aval (offre locale). Par conséquent, le modèle de transmission de cellule est tout à fait indiqué pour l'implémentation en parallèle. Le désavantage causé par la restriction est que les carrefours ayant de multiples arcs dans l'étoile sortante et l'étoile entrante doivent être représentés comme des échangeurs ayant une bifurcation à chaque abord de carrefour afin de diviser le courant de trafic.

L'évaluation du modèle de transmission de cellule est à la fois qualitative et quantitative. Le programme CellNetLoad a été conçu en un seul fil (*thread*) et propose une simulation à base d'événements. Le flot est propagé selon la méthode de transmission de cellule, les feux sont modélisés par phases explicites, et les carrefours ne sont pas limités aux carrefours de convergence ou de bifurcation. Une série de réseaux d'essai a été mise au point pour vérifier l'exactitude des résultats de la simulation. De petits réseaux ont été utilisés pour isoler des caractéristiques particulières des réseaux de trafic tandis qu'un réseau d'essai plus complexe a été utilisé pour évaluer la performance générale du modèle de transmission de cellule. L'analyse de l'impact de la discrétisation de l'espace-temps sur les résultats de la simulation a, entre autres, été réalisée. Les temps d'exécution de la simulation ont été consignés pour fins de comparaison.

L'une des conclusions de l'évaluation est que CellNetLoad a une complexité quadratique à l'inverse de l'intervalle de temps.

Pour établir la fidélité de CellNetLoad, un repère a été utilisé qui provenait d'un programme de simulation microscopique du trafic, appelé INTEGRATION. Il est permis de remettre en question la fidélité avec laquelle INTEGRATION représente la réalité, mais le programme permet une modélisation du trafic plus détaillée que le modèle de transmission par cellule. Par conséquent, il doit certainement se rapprocher davantage de la réalité. Il y a des avantages indéniables à l'utilisation d'une simulation microscopique en tant que repère au lieu de l'observation d'un véritable réseau de trafic. Le réseau d'essai peut être adapté sur mesure pour isoler certains éléments, le nombre de variables stochastiques peut être modifié et la simulation peut être recommencée.

Les résultats de la simulation avec CellNetLoad soutiennent la comparaison avec INTEGRATION sur les réseaux où l'effet des changements de voie est négligeable. Les ondes de choc convergent vers la solution du modèle LWR. Les divergences entre les résultats de la simulation avec CellNetLoad et le repère sont imputables à certaines caractéristiques du programme ou à l'absence de celles-ci. D'abord, la relation flot-densité linéaire par morceaux ne permet pas la diffusion des éventails et le flot non congestionné se déplace à la vitesse à écoulement libre. Ensuite, les effets des changements de voie ne peuvent être mesurés qu'en utilisant un programme extérieur. La convergence asymétrique des voies causée par la disparition d'une voie et le tricotage en amont en raison des bifurcations n'est pas reproduite par le modèle de transmission de cellule. Enfin, la logique du premier entré premier sorti est trop contraignante lorsqu'il y a des bifurcations à plusieurs voies.

La modélisation des intersections présente de nombreuses difficultés : la réduction de la capacité causée par les mouvements opposés doit être définie à l'extérieur ; les délais aléatoires, les délais causés par les changements de voie et les délais causés par les mouvements opposés ne sont pas reconnus ; de plus, les feux doivent avoir des phases explicites afin de pouvoir inclure l'attente cyclique et les délais qui en découlent.

Compte tenu des résultats de l'évaluation, certaines améliorations ont été apportées au modèle de transmission de cellule. Les améliorations ont ensuite été testées sur les mêmes réseaux. Voici quelques améliorations concernant la propagation du flot : une relation flot-densité non linéaire, la compensation de l'erreur dans la capacité d'entreposage des arcs et la discrétisation du flot. Voici les améliorations au modèle des carrefours : l'assouplissement de la contrainte qui ne permettait que des carrefours de bifurcation et de convergence, l'assouplissement de la discipline du premier entré premier sorti et la modélisation explicite du marquage des voies aux abords des carrefours.

L'utilisation de la relation flot-densité non linéaire dans le modèle de transmission de cellule a pour effet d'augmenter la dissipation des ondes de choc, mais de diminuer les temps de déplacement. Par ailleurs, les éventails sont modélisés de façon plus réaliste. La modélisation plus précise de la capacité d'entreposage des arcs produit un débordement plus opportun. La discrétisation du flot introduit un biais qui a pour effet de limiter le flot. Le biais en question s'aggrave avec la grandeur de l'unité du flot discrète, avec la diminution de l'intervalle de temps et à mesure que le flot s'approche de la capacité. En raison de ce biais, CellNetLoad ne converge pas vers la solution du modèle hydrodynamique LWR.

Le modèle généralisé des carrefours, dans CellNetLoad, représente les carrefours de façon géométriquement réaliste. Pour y arriver, cinq algorithmes sont proposés. L'algorithme de base sert à l'application de la discipline du premier entré premier sorti, dont l'effet est de permettre les dépassements des paquets, mais seulement à l'intérieur d'une même cohorte. Par conséquent, les débordements gênent tous les courants de trafic dès qu'un courant de trafic déborde dans l'abord

du carrefour. La capacité de sortie et l'offre locale sont réparties sur les différents courants de trafic en fonction de la demande partielle. L'algorithme de base convient pour les abords de carrefour à voie unique. L'algorithme relaxé de premier entré premier sorti de cohorte permet les dépassements des paquets dans la cohorte (ou les cohortes) qui se trouve(nt) en tête de la file d'attente représentant la demande locale. Par conséquent, il y a un délai dans le débordement. L'algorithme de cellule de premier entré premier sorti permet les dépassements dans la dernière cellule de l'abord du carrefour, ce qui a pour effet de retarder davantage le débordement. Les voies explicites avec permissions sont modélisées dans l'algorithme relaxé de premier entré premier sorti de voie et dans l'algorithme de premier entré premier sorti de cellule de voie.

Il n'existe aucun mécanisme dans le modèle de transmission de cellule pour rendre les effets de l'espacement entre les véhicules là où les mouvements opposés sont permis ou pour rendre les effets du tricotage à un abord de bifurcation. Par conséquent, la réduction de la capacité doit être définie à la mise à exécution du modèle. La modélisation explicite des voies permet de définir la réduction de la capacité de celles-ci. Toutefois, établir la réduction de la capacité temporelle par voie n'est pas chose facile.

Les algorithmes de carrefours sans voies explicites convergent en un nombre d'itérations qui est majoré par le nombre de courants de trafic avec demande positive. Les algorithmes ayant des voies explicites convergent, mais le nombre d'itérations est non borné. Durant les périodes d'encombrement, les algorithmes de cellule de premier entré premier sorti sont grandement ralentis. Par conséquent, les temps d'exécution de la simulation sont plus longs que ceux de la simulation microscopique qui a été utilisée comme repère.

Le réseau de corridor a confirmé que le modèle de carrefour est essentiel au réalisme de la simulation. L'absence de changement de voie et l'erreur dans la modélisation des dépassements et des débordements ont eu des effets considérables sur les résultats de la simulation.

Mots-clefs : Modèle macroscopique d'affectation dynamique

Executive Summary

The first objective of this research is to provide a comprehensive explanation of the theory behind the cell-transmission model as a solution method for the dynamic network loading (DNL) problem. The cell-transmission model was developed with contributions from diverse fields. This has led to gaps in the literature, and alterations of the model which contradict fundamental assumptions in the theory.

The second objective of this thesis is to evaluate the accuracy of the cell-transmission model, and to investigate enhancements which may improve the accuracy. Of course, the implications of these enhancements on the simulation execution speed are considered, but it is of secondary importance in this work.

The literature related to the cell-transmission model was thoroughly investigated. The cell-transmission model is quite recent, first defined as a solution method for the dynamic network loading problem for traffic networks by C.F. Daganzo in 1994. However, the cell-transmission model draws on early findings from many fields of study, such as traffic engineering (Greenshields, 1935), traffic flow theory (Lighthill and Whitham, 1955; Richards 1956), numerical methods (Lax, 1954), and fluid dynamics (Godunov, 1959). Substantial contributions were also made by Ansorge (1990) and Lebacque (1996).

Context is provided by formally describing the DNL problem, and classifying the alternative solution methods. Several background topics are essential to the development of the cell-transmission model. The state variables, namely flow, density, and speed, are defined along with their fundamental relationship. The flow-density relationship is examined in more detail. Examples of non-linear, multi-regime, and piecewise linear flow-density relationships are provided. Finally, the law of conservation of flow is developed.

The LWR hydrodynamic model (Lighthill and Whitham, 1955; Richards 1956) is defined using the conservation of flow and an equilibrium flow-density relationship as a first order model derived from fluid dynamics. The LWR model is a macroscopic solution method to the DNL problem. An initial value problem, in which the initial density is known for a roadway with an equilibrium flow-density relationship, can be solved using the weak formulation of the LWR model. However, the solution is not unique, so the entropy condition is introduced to select the physically relevant solution. The exact solution to the initial value problem may be found by an analytical method, which is called the method of characteristics. Once density and flows are known, the fundamental relationship can be used to determine the speed, and thus travel time.

Finite difference approximations to the LWR model are investigated for the purpose of developing a more easily mechanised solution method for the DNL problem. Lax (1954) first proposed a crude differencing scheme. Godunov (1959) developed an approximate solution, which was adapted to traffic networks by Lebacque (1996) The initial value problem is essentially reduced to a series of generalised Riemann problems. Lebacque (1996) drew the analogy between the Godunov method and the concepts of local demand and local supply.

Daganzo (1994) found an even simpler solution to the generalised Riemann problem as the minimum of the local demand and the local supply. A trapezoidal function is used for the equilibrium flow-density relationship. Time is discretised into a series of equal length time steps. Each arc is represented by a series of cells, each with length defined as the distance travelled during one time step by a vehicle moving at the free speed of that arc. It is shown that the solution to the cell-transmission model converges to the solution of the LWR model for finer time-space discretisations.

The cell-transmission model also has a node model. Only merge and diverge intersections may be considered using a flow maximising logic. A result of this restriction is that the flow out of each cell depends only on the density of the current cell (local demand), and the density of the downstream cell(s) (local supply). Therefore, the cell-transmission is well suited to a parallel implementation. The drawback of this restriction is that intersections with multiple arcs in both the forward star and the backward star must be represented as an interchange with a diverge on each intersection approach to split the streams.

The cell-transmission model is evaluated qualitatively and quantitatively. CellNetLoad was developed as a single thread, event-based simulation. Flow is propagated according to the cell-transmission method; signals are modelled with explicit phasing; and intersections are *not* restricted to merges and diverges. A set of test networks were developed to determine the accuracy of the simulation results. Small networks were used to isolate particular network features, and a more complex test network was then used to evaluate the performance of the cell-transmission model on a more realistic network. The evaluation includes an analysis of the impact of the discretisation of time-space on the simulation results. The simulation execution times were recorded for comparison. It was determined that CellNetLoad has quadratic complexity with the inverse of the time step.

The accuracy of CellNetLoad was determined by comparison with a benchmark. The benchmark was derived from a microscopic traffic simulation program, called INTEGRATION. Though the realism of INTEGRATION is questionable, it models traffic in greater detail than the cell-transmission model. So, it is reasonable to assume that it is closer to reality. There are several advantages of using a microscopic simulation as the benchmark instead of observations of an existing traffic network. The test network can be customised and altered to isolate certain features, the degree of stochasticity can be controlled, and the simulation is repeatable.

The simulation results compared favourably with INTEGRATION for networks in which the effect of lane changing was not significant. Shockwaves converges to the solution of the LWR model. Several features, or lack thereof, are responsible for the discrepancy between the simulation results of CellNetLoad and the benchmark. Firstly, the piecewise linear equilibrium flow-density relationship prevented spreading of fans, and all uncongested flow travelled at the free speed. Secondly, the effects of lane-changing can only be included with external models. The asymmetric merging of lane flows due to a lane-drop and weaving upstream of diverges is not captured by the cell-transmission model. Thirdly, first-in-first-out logic is too restrictive for multi-lane diverges.

The modelling of intersections has a host of problems. The capacity reduction due to opposed movements must be specified externally. Random delay, delay due to lane-changing, and delay to opposed movements are also ignored. Explicit signal phasing is necessary for the inclusion of cyclic queuing and the associated delay.

Based on the findings of the evaluation, several enhancements to the cell-transmission were implemented, then evaluated on the same test networks. The enhancements to flow propagation include a non-linear equilibrium flow-density relationship, compensation for the error in the arc storage capacity, and the discretisation of flow. The enhancements to the intersection model involve the relaxation of the constraint that permits only diverge and merge intersections, the relaxation of FIFO discipline, and the explicit modelling of lane striping at intersection approaches.

The use of a non-linear equilibrium flow-density relationship with the cell-transmission model tends to increase the dissipation of the shockwave, but the uncongested travel times are improved, and fans are modelled more realistically. Correctly modelling the arc storage capacity produces more timely spillback. The discretisation of flow introduced a bias to restrain flow, which worsens with increasing vehicle unit size, with decreasing time step, and as the flow approaches the capacity. Because of this bias, CellNetLoad will not converge to the solution of the LWR hydrodynamic model.

The generalised intersection model of CellNetLoad allows realistic representations of intersections. Five algorithms for the intersection model are available in CellNetLoad. The basic algorithm enforces cohort FIFO discipline, which allows overtaking of packets within the same cohort only. As a result, spillover impedes all streams as soon as one stream spills back on the approach. The discharge capacity and local supply are distributed to competing streams in proportion to the partial demand. This algorithm is appropriate for single-lane approaches. The relaxed cohort FIFO algorithm permits overtaking among the packets in the cohort(s) at the front of the cohort queue which represent the local demand, which delays spillover. The cell FIFO algorithm permits overtaking within the last cell of the approach, further delaying spillover. Explicit lanes with permissions are modelled in the lane relaxed FIFO algorithm and the lane cell FIFO algorithm.

There is no mechanism in the cell-transmission model to capture the effects of gap acceptance for an opposed permissive movement, or the effects of weaving on a diverge approach. Therefore, the capacity reduction must be specified in the model input. Modelling explicit lanes permitted the specification of lane capacity reductions. However, determining the temporal, lane-specific capacity reductions is not trivial.

The intersection algorithms without explicit lanes converge in a number of iterations that is bounded from above by the number of streams with positive demand. The algorithms with explicit lanes converge, but the number of iterations is unbounded. During heavy congestion, the cell FIFO algorithms are slowed considerably, resulting in simulation execution times far longer than even the microscopic simulation used as the benchmark.

The corridor network confirmed that the intersection model is crucial to the accuracy of the simulation. The lack of lane changing, and the error in modelling overtaking and spillback were found to drastically effect the simulation results for the entire network.

Table of Contents

- **hSommaire** ii
- **Abstract** iv
- **Résumé** vi
- **Executive Summary** vi
- **Table of Contents** xviii
- **List of Figures** xxii
- **List of Tables** xxvi
- **Notation** xxvii
- **Glossary** xxxi
- **Acknowledgements** xxxv
- **Chapter 1 Introduction** 1
  - 1.1 Objectives of research 1
  - 1.2 Approach methodology 1
  - 1.3 Structure of the thesis 3
- **Chapter 2 Theoretical Foundation and Qualitative Evaluation of the Cell-transmission Model** 5
  - 2.1 Introduction 5
  - 2.2 Approaches to Dynamic Network Loading 6
    - 2.2.1 Analytical dynamic network loading models 9
    - 2.2.2 Macroscopic dynamic network loading models 11
    - 2.2.3 Microscopic dynamic network loading models 12
    - 2.2.4 Mesoscopic dynamic network loading models 13
  - 2.3 Development of the Cell-transmission Model 15
    - 2.3.1 Speed-flow-density relationship 16
    - 2.3.2 Equilibrium flow-density curves 17
      - 2.3.2.1 Non-linear equilibrium flow-density relationships 20
      - 2.3.2.2 Linear equilibrium flow-density relationships 22
    - 2.3.3 Conservation law 24
    - 2.3.4 Lighthill-Whitham-Richards hydrodynamic model 25
    - 2.3.5 Kinematic waves 26
    - 2.3.6 Shockwaves at density discontinuities 29
    - 2.3.7 Analytical solution to the LWR hydrodynamic model 34
      - 2.3.7.1 Smooth and weak solutions 34
      - 2.3.7.2 The entropy condition and driver's ride impulse 37
      - 2.3.7.3 Unique solution to the LWR hydrodynamic model 41
      - 2.3.7.4 Solution by the method of characteristics 44
    - 2.3.8 Finite difference approximations to the LWR hydrodynamic model 45
      - 2.3.8.1 Numerical methods 45
      - 2.3.8.2 Godunov approximate solution 48
  - 2.4 The Cell-transmission Model 56
    - 2.4.1 Formulation of the cell-transmission model 56

2.4.2 Error quantification ... 57
    2.4.2.1 Error along a characteristic ... 58
    2.4.2.2 Shockwave representation ... 60
2.4.3 Approximation of the solution of the LWR model ... 68
2.4.4 Adaptation to traffic networks ... 69
    2.4.4.1 Simple intersection models ... 69
    2.4.3.2 Origins and destinations ... 77
    2.4.3.3 Multi-commodity flow ... 78
2.5 Discussion of the Cell-transmission Model ... 80
    2.5.1 Theory and methodology ... 81
    2.5.2 Policy sensitivity ... 85
    2.5.3 Simplicity ... 86
    2.5.4 Data requirements ... 87
    2.5.5 Domain of application ... 88
    2.5.6 Implications of theory on practical applications ... 89

Chapter 3 Implementation of the Cell-transmission Model 93

3.1 Introduction ... 93
3.2. Development context ... 94
3.3 Network representation ... 95
3.4 Input-output structure ... 98
3.5 CellNetLoad simulation program ... 102
    3.5.1 Initialisation ... 103
    3.5.2 Simulation ... 103
        3.5.2.1 Route event ... 103
        3.5.2.2 Demand event ... 104
        3.5.2.3 Signal event ... 105
        3.5.2.4 Incident event ... 107
        3.5.2.5 Clock tick event ... 108
        3.5.2.6 Output statistics event ... 118
        3.5.2.7 Simulation terminate event ... 118
    3.5.3 Post-processing of output ... 118
3.6 Conclusions ... 120

Chapter 4 Quantitative Evaluation of the Cell-transmission Model 121

4.1 Introduction ... 121
4.2 Evaluation strategy ... 122
4.3 Linear network ... 124
    4.3.1 INTEGRATION simulation results ... 126
    4.3.2 CellNetLoad simulation results ... 138
4.4 Diverge ... 148
    4.4.1 INTEGRATION simulation results ... 150
    4.4.2 CellNetLoad simulation results ... 152
4.5 Merge ... 154
    4.5.1 INTEGRATION simulation results ... 156
    4.5.2 CellNetLoad simulation results ... 159
4.6 Intersection ... 161
    4.6.1 INTEGRATION simulation results ... 163
    4.6.2 CellNetLoad simulation results ... 167
4.7 Conclusions ... 172

Chapter 5 Extensions to Flow Propagation 174

5.1 Introduction 174
5.2 Non-linear equilibrium flow-density relationship 175
  5.2.1 Implementation 175
  5.2.2 Effect of the non-linear equilibrium flow-density relationship 178
5.3 Arc discretisation 185
  5.3.1 Error in arc length 185
  5.3.2 Reduction in time step due to intersection model 187
  5.3.3 Variable time step 188
5.4 Discretisation of flow 189
  5.4.1 Implementation 189
  5.4.2 Effect of the discretisation of flow 191
5.3 Computational efficiency 197
  5.3.1 Analysis of complexity 197
  5.3.2 Comparison of execution times for active enhancements in CellNetLoad 202
5.4 Conclusions 204

Chapter 6 Intersection Model 207

6.1 Introduction 207
6.2 The basic intersection model 208
  6.2.1 Discussion of weaknesses 210
  6.2.2 Alternative intersection representation 212
    6.2.2.1 Graph representation and the max flow problem 212
    6.2.2.2 Convergence of the generalised intersection algorithm 214
    6.2.2.3 Simulation results for the alternative intersection representation 218
6.3 Relaxation of cohort FIFO discipline 221
  6.3.1 Implementation 222
  6.3.2 Effects on the diverge network 226
  6.3.3 Effects on the merge network 229
  6.3.4 Effects on the intersection network 230
6.3 Explicit lanes 230
  6.3.1 Implementation 231
  6.4.2 Effects on the diverge network 239
  6.4.3 Effects on the merge network 242
  6.4.4 Effects on the intersection network 242
6.5 Effects on the simulation execution time 246
  6.5.1 Simulation execution time of the diverge network 248
  6.5.2 Simulation execution time of the merge network 250
  6.5.2 Simulation execution time of the intersection network 252
6.6 Conclusions 256

Chapter 7 Simulation of a Commuter Corridor 259

7.1 Introduction 259
7.2 Corridor network description 260
7.3 Simulation results 261
  7.3.1 INTEGRATION simulation results 262
  7.3.2 CellNetLoad simulation results 273
    7.3.2.1 Effect of time discretisation 274
    7.3.2.2 Effect of the enhanced intersection algorithms 279
    7.3.2.3 Effect of the enhanced flow propagation 283
7.4 Simulation execution times 288
7.5 Discussion 290

Chapter 8 Conclusions and Recommendations 292
8.1 Contributions to theory ... 292
8.2 Findings and conclusions ... 296
8.3 Recommendations ... 298
References 300
Appendix. Format of Input Files for INTEGRATION I
VITA VIII

List of Figures

Figure 2.1 Greenshields speed-flow-density relationship 18
Figure 2.2 Generalised equilibrium flow-density curve 19
Figure 2.3 Non-linear equilibrium flow-density relationships 21
Figure 2.4 Linear equilibrium flow-density relationships 22
Figure 2.5 A shockwave at a density discontinuity 31
Figure 2.6 Relationship between the flow-density curve and the position-time graph for a discontinuous density along a homogeneous road 32
Figure 2.7 Two weak solutions of the LWR hydrodynamic model for a density discontinuity36
Figure 2.8 Graphical interpretation of the entropy condition and driver's ride impulse 40
Figure 2.9 Shockwaves and fans for a triangular equilibrium flow-density relationship 42
Figure 2.10 Density profile for an initial density discontinuity 43
Figure 2.11 Density at lattice points for an instant in time 46
Figure 2.12 Notation for space discretisation and cell placement 49
Figure 2.13 General solution algorithm of the Godunov scheme 49
Figure 2.14 Modified solution algorithm of the Godunov scheme 50
Figure 2.15 Generalised Riemann problem with discontinuous equilibrium flow-density at cell boundary 52
Figure 2.16 Traffic supply and traffic demand curves at a cell boundary 55
Figure 2.17 Influence diagrams for shocks with gradual density transitions 65
Figure 2.18 Influence diagrams for shocks with abrupt and gradual density transitions 66
Figure 2.19 Simple intersections 70
Figure 2.20 Max flow graph for a merge node 73
Figure 2.21 Max flow graph for a diverge node 75
Figure 2.22 Max flow graph for a single-lane diverge node 76
Figure 2.23 Comparison of flow-density relationships 82
Figure 3.1 Piecewise linear equilibrium flow-density relationship used in CellNetLoad .. 100
Figure 3.2 Process method of a demand event 105
Figure 3.3 Process method of a signal timing event 106
Figure 3.4 Process method of a signal phase event 108
Figure 3.5 Prepare cohorts to depart from all last cells during a clock tick event 109
Figure 3.6 Prepare cohorts to depart an origin 111
Figure 3.7 Cohort FIFO algorithm
- overtaking is permitted for packets in the same cohort 113
Figure 4.1 Linear test network 125
Figure 4.2 Density from INTEGRATION on the linear test network with 25-m arcs 127
Figure 4.3 Flow from INTEGRATION at the lane drop of the linear test network 129
Figure 4.4 Density from INTEGRATION on the linear test network with 50-m arcs 132
Figure 4.5 Density from INTEGRATION at fixed points on the linear test network with 125-m arcs 134
Figure 4.6 Density from INTEGRATION at fixed points on the linear test network with 25-m arcs 136
Figure 4.7 Path travel times from INTEGRATION on the linear test network 138
Figure 4.8 Density from CellNetLoad on the linear test network with 25-m cells (1-s time step)
.... 139
Figure 4.9 Density from CellNetLoad on the linear test network with 50-m cells (2-s time step)

.... 141

Figure 4.10 **Density from CellNetLoad at** fixed positions on the linear test network with 25-m cells (1-s time step) 142

Figure 4.11 **Density from CellNetLoad at fixed** points on the linear test network with 125-m cells (5-s time step) 143

Figure 4.12 **Path travel times from CellNetLoad** on the linear test network 144

Figure 4.13 **Density from CellNetLoad on the** linear test network with 25-m arcs (1-s time step) and a reduced capacity at the lane-drop 146

Figure 4.14 **Density from CellNetLoad on the** linear test network with 50-m arcs (2-s time step) and a reduced capacity at the lane-drop 147

Figure 4.15 **Path travel times from CellNetLoad** on the linear test network with a reduced capacity (for time steps 1s and 2s)at the lane-drop 148

Figure 4.16 **Diverge test network** 149

Figure 4.17 **Flow from INTEGRATION at a** diverge 151

Figure 4.18 **Flow from CellNetLoad at a** diverge 153

Figure 4.19 **Merge test network** 155

Figure 4.20 **Flow from INTEGRATION at a** merge with constant departure headways 157

Figure 4.21 **Flow from INTEGRATION at a** merge with exponentially distributed departure headways 159

Figure 4.22 **Flow from CellNetLoad at a** merge 161

Figure 4.23 **Intersection test network** 162

Figure 4.24 **Intersection test networkSimulation results from INTEGRATION of** an intersection 165

Figure 4.25 **Interchange representation of an intersection** by the traditional cell-transmission model 167

Figure 4.26 **Simulation results from CellNetLoad of** an intersection with interchange representation and static signals 169

Figure 4.27 **Simulation results from CellNetLoad of** an intersection with interchange representation and static signals compared to INTEGRATION 170

Figure 4.28 **Simulation results from CellNetLoad of** an intersection with interchange representation and signal phasing 171

Figure 4.29 **Simulation results from CellNetLoad of** an intersection with interchange representation and signal phasing compared to INTEGRATION 172

Figure 5.1 **Density from CellNetLoad on the** linear network with 25-m cells (1-s time step) and a non-linear equilibrium flow-density relationship 179

Figure 5.2 **Density from CellNetLoad at fixed** positions on the linear network with 25-m arcs/cells (1-s time step) and a non-linear equilibrium flow-density relationship . 182

Figure 5.3 **Density from CellNetLoad at fixed** positions on the linear network with 125-m arcs/cells (5-s time step) and a non-linear equilibrium flow-density relationship . 183

Figure 5.4 **Path travel times from CellNetLoad** on the linear network with a non-linear equilibrium flow-density relationship 184

Figure 5.5 **Density from CellNetLoad on the** linear network with 125-m cells (5-s time step) and probabilistic rounding to integer vehicle units 192

Figure 5.6 **Density from CellNetLoad on the** linear network with 125-m arcs/cells (5-s time step) and probabilistic rounding to integer vehicle units 194

Figure 5.7 **Density from CellNetLoad on the** linear network with 25-m arcs/cells (1-s time step) and probabilistic rounding to integer vehicle units 195

Figure 5.8 **Path travel times from CellNetLoad** on the linear network with probabilistic rounding to integer vehicle units 196

Figure 5.9 **Cumulative execution time of the** clock tick events for a linear network 199

Figure 5.10 **Cumulative execution time of** the output statistics events for a linear network .. 200

Figure 5.11 **Remaining execution time of** the simulation of a linear network 201
Figure 5.12 **Execution time and simulation** time of CellNetLoad on the linear network 202
Figure 5.13 **Execution time and simulation** time of CellNetLoad on the linear network with log-scaling 203
Figure 6.1 Graph representation of a 4-leg intersection with the interchange representation . 209
Figure 6.2 Graph representation of a 4-leg intersection 213
Figure 6.3 Example of an intersection for which the generalised intersection algorithm does not find the max flow solution when the distribution of the discharge capacity and local supply are relaxed 215
Figure 6.4 Simulation results from CellNetLoad of an intersection 219
Figure 6.5 Simulation results from CellNetLoad of an intersection with reduced capacity 220
Figure 6.6 Simulation results from CellNetLoad of an intersection with reduced capacity compared to INTEGRATION 221
Figure 6.7 Cell FIFO algorithm for the determination of intersection flow
– overtaking permitted among packets in the same cell 223
Figure 6.8 Relaxed cohort FIFO algorithm for the determination of intersection flow
– overtaking permitted among the packets at the front of the cohort queue which represent the local demand 225
Figure 6.9 Flow from CellNetLoad at a diverge with cell FIFO discipline 227
Figure 6.10 Flow from CellNetLoad at a diverge with relaxed cohort FIFO discipline 228
Figure 6.11 Flow from CellNetLoad at a merge with cell FIFO discipline 229
Figure 6.12 Cell FIFO algorithm for the determination of intersection flow
– overtaking permitted among packets in the same cell
– explicit lanes with lane striping permission on each approach 232
Figure 6.13 Relaxed cohort FIFO algorithm for the determination of intersection flow
– overtaking permitted among the packets at the front of the cohort queue which represent the local demand
– explicit lanes with lane striping permission on each approach 237
Figure 6.14 Flow from CellNetLoad at a diverge with cell FIFO discipline and explicit lanes 240
Figure 6.15 Flow from CellNetLoad at a diverge with relaxed cohort FIFO discipline and explicit lanes 241
Figure 6.16 Flow from CellNetLoad at a merge with explicit lanes 243
Figure 6.17 Lane capacity factors for the intersection network 243
Figure 6.18 Simulation results from CellNetLoad of an intersection with cell FIFO discipline and explicit lanes 244
Figure 6.19 Simulation results from CellNetLoad of an intersection with cell FIFO discipline and explicit lanes compared to INTEGRATION 245
Figure 6.20 Simulation results from CellNetLoad of an intersection with relaxed cohort FIFO discipline and explicit lanes 247
Figure 6.21 Simulation execution time of CellNetLoad for a diverge network 249
Figure 6.22 Simulation execution time of CellNetLoad for a merge network 251
Figure 6.23 Simulation execution time of CellNetLoad for the intersection network with light flow (linear scale) 253
Figure 6.24 Simulation execution time of CellNetLoad for the intersection network with light flow (log scale) 254
Figure 6.25 Simulation execution time of CellNetLoad for the intersection network with heavy flow 255
Figure 6.26 Simulation execution time of CellNetLoad for the intersection network with heavy flow and an incident 256
Figure 7.1 **Corridor network** 260
Figure 7.2 **Simulation results for arc 1** of the corridor network 263

Figure 7.3 **Simulation results for arc 2** of the corridor network 264
Figure 7.4 **Simulation results for arc 3** of the corridor network 265
Figure 7.5 **Simulation results for arc 4** of the corridor network 266
Figure 7.6 **Simulation results for arc 10** of the corridor network 267
Figure 7.7 **Simulation results for arc 11** of the corridor network 268
Figure 7.8 **Temporal path travel times of** the corridor network 269
Figure 7.9 **Simulation results for arc 11** of the corridor network with relaxed FIFO discipline281
Figure 7.10 **Simulation results for arc 3** of the corridor network with cell FIFO discipline and enhanced flow propagation 284
Figure 7.11 **Temporal path travel times of** the corridor network with cell FIFO and enhanced flow propagation 286
Figure 7.12 **Simulation results for arc 1** of the corridor network with cell FIFO discipline and enhanced flow propagation 287
Figure 7.13 **Simulation execution time of CellNetLoad** for the corridor network....289

List of Tables

Table 2.1 **Solution to generalised Riemann problem** $q(x_i, t_j)$ by exhaustive enumeration53
Table 3.1 **Activation of enhanced features of** CellNetLoad using the command line flags 102
Table 7.1 **Arc discretisation and error in** arc length 275

Notation

G	graph, $G = < N, A >$
$\widetilde{G}$	network, i.e. graph with node and arc attributes, $\widetilde{G} = < N, A >$
N	set of nodes in graph G
n	node, $n \in N$
A	set of arcs in graph G
a	arc, $a = \left(n_a^-, n_a^+\right) \in A \subset N \times N$
A_n^+	forward star of node $n \in N$, $A_n^+ \subset A$
A_n^-	backward star of node $n \in N$, $A_n^- \subset A$
O	set of origin nodes, $O \subset N$
D	set of destination nodes, $D \subset N$
o	origin node, $o \in O \subset N$
d	destination node, $d \in D \subset N$
p	origin-destination pair, $p = (o, d) \in O \times D$
R_p	set of routes/paths for origin-destination pair p

r	route, $\mathrm{r} = \{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_{\lvert \mathrm{r} \rvert}\} \in \mathrm{R}_\mathrm{p}$ where arcs $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_{\lvert \mathrm{r} \rvert} \in \mathrm{A}, \mathrm{a}_1 \in \mathrm{O} \times \mathrm{N}, \mathrm{a}_{\lvert \mathrm{r} \rvert} \in \mathrm{N} \times \mathrm{D}$
i, j	cell
ℓ	lane
t	time, $t \in [0, \bar{t}]$
Δt	simulation time step
$\bar{t}$	time of end of simulation
$g_\mathrm{r}(t)$	demand rate on route $\mathrm{r} \in \mathrm{R}_\mathrm{p}$ at time t
$\mathrm{T}(t)$	set of travel times at time t
$T_\mathrm{a}(t)$	travel time $T_\mathrm{a}(t) \in \mathrm{T}(t)$ for a vehicle which enters arc $\mathrm{a} \in \mathrm{A}$ at time t
x	position $x \in [0, \bar{x}]$ on an arc, starting from the upstream end
$\bar{x}$	length of an arc
Δx_i	length of cell i for a discretised arc
l	subscript meaning left, e.g. $x_l = \lim_{\varepsilon \to 0}(x - \varepsilon)$
r	subscript meaning right, e.g. $x_r = \lim_{\varepsilon \to 0}(x + \varepsilon)$

q	flow
$q_{\max}$	maximum flow or capacity flow or saturation flow rate
$\tilde{q}$	equilibrium flow
v	speed
v_f	free-flow speed
v_c	speed-at-capacity
k	density
k_{jam}	jam density
k_c	density-at-capacity or critical density
w	kinematic wave speed
$w_{\max}$	maximum backward kinematic wave speed
u	shockwave speed
n	number of vehicles
$n_{\mathrm{ij}\tau}^{\mathrm{d}}(t)$	number of vehicles present in cell i at time t which entered the cell during the arrival time interval $[\tau, \tau+\Delta t)$ and going to destination $\mathrm{d} \in \mathrm{D}$ via downstream cell j
Δn	vehicle discretisation unit

Q_{i}	number of vehicles which may enter or exit cell i at the capacity flow $q_{\max}$ during one simulation time step Δt, $Q_{\mathrm{i}} = q_{\max} \cdot \Delta t$
L_{i}	set of lanes in cell i
N_{i}	number of vehicles permitted in cell i at jam density k_j, $N_{\mathrm{i}} = k_j \cdot \Delta x_{\mathrm{i}} \cdot \left\Vert \mathrm{L}_{\mathrm{i}} \right\Vert$
$y_{\mathrm{ij}\ell}(t)$	number of vehicles which exit lane ℓ of cell i and enter cell j during the time interval $[t, t + \Delta t)$
$\alpha_{\mathrm{ij}}^{m}(t)$	local demand coefficient, i.e. the proportion of local demand in upstream cell i allocated to downstream cell j during the time interval $[t, t + \Delta t)$ at iteration m of the node solution algorithm
$\beta_{\mathrm{ij}}^{m}(t)$	local supply coefficient, i.e. the proportion of local supply in downstream cell j allocated to upstream cell i during the time interval $[t, t + \Delta t)$ at iteration m of the node solution algorithm
D	local demand
S	local supply

Glossary

arc	the roadway which connects two nodes in a graph of a traffic network
backward star	the arcs which enter a particular node
car-following	the behaviour of a following vehicle as a reaction to the preceding vehicle in the same lane
cell	a discretisation of the arc length
cycle length	the duration of one cycle of a signal, which includes each phase once
density-at-capacity	the number of vehicles in a given length of a lane when the vehicles are travelling at the saturation flow rate
discharge capacity	the capacity at the exit of an arc
diverge	a node with a single arc in the backward star and multiple arcs in the forward star
DNL	dynamic network loading
FIFO	first-in-first-out
flow capacity	the maximum possible flow, accounting for capacity reducing phenomena, such as traffic control devices, incidents, weaving, and vehicle conflicts

flow state	a region in space time with homogeneous state variables: flow, density, and speed
forward star	the arcs which exit a particular node
free speed	the maximum speed of a vehicle on an approach, which occurs when the vehicle has an infinite headway
green time	the sum of all intervals which indicate green for a particular approach to a signalised intersection
h	hour
headway	the time or distance, between two successive vehicles in the same lane, measured from the same point on each vehicle
INTEGRATION	a microscopic traffic network simulation model
intersection	a node with multiple arcs in the backward star and/or multiple arcs in the forward star
interval (of a signal)	the period of time which has a single colour, either red, green, or yellow, for a signal
jam density	the number of vehicles in a given length of a lane when the vehicles are stopped
km	kilometre
lane striping	the turning permissions of each lane on an intersection approach

local demand	the quantity of vehicles (per unit time) which is prepared to move downstream based on upstream conditions
local supply	the quantity of vehicles (per unit time) which are permitted to move downstream based on downstream conditions
m	metre
merge	a node with multiple arcs in the backward star and a single arc in the forward star
phase (of a signal)	the period of time which assigns priority to one-or more intersection approaches at a signalised intersection
s	second
saturation flow rate	the maximum flow on an arc excluding capacity reduction
shockwave	the boundary between two flow states
signal timing plan	the offset, the duration of each phase, and the duration of each interval of a signal
speed-at-capacity	the speed of a vehicle on an approach which has a volume equal to the saturation flow rate
storage capacity	the maximum possible storage for a given length of an arc, accounting for capacity reducing phenomena, such as incidents
time step	discretisation of time; the time interval used to step through a simulation

vph	vehicles per hour
vphpl	vehicles per hour per lane

Acknowledgements

Thank you...

to Mike Florian for giving me a wide berth when I wanted it, giving me a push when I needed it, and giving generously of your support always. You deserve your reputation as a scholar and a gentleman.

to the other members of the review committee, Jean-Yves Potvin, Jaime Barceló, Michel Gendreau, and Michel Delfour for your thorough assessment and your constructive criticism.

to NSERC for providing me with an income while pursuing my studies. I hope you got your money's worth.

to Mike Mahut, Karim Er-Rafia, and Vittorio Astarita, who rounded-out our research team. I couldn't have done it without you, because I probably wouldn't have understood what I was doing.

to the administrative staff of the CRT Clairette Simard, Murielle Deslile, Claudine St-Pierre, and Martine Gemme, for your professionalism and your smiles.

to the tech support, Guy Bordeleau, Daniel Charbonneau, and Luc Rocheleau, who always seem to be available to answer my questions and to fix my blunders.

to the staff at the bibliothèque for your books and your help locating them.

to Kim Kilpatrick for your timely effort on the French translations.

to Jill Gasco, who I will be proud to spend the rest of my life with, for your love and support. I didn't know people were made so sweet until I met you.

to my family for their support through all my ups and downs.

to my good friends, Donagh, Mark, Christian, Marty, Chris, and Ray, who had open ears and cold beers. Now I will have to find another topic to bore you.

to Montréal for building a world class learning centre right in my home town.

to Starbucks Coffee for locating a franchise only a short walk from school.

I began my doctoral studies with my eyes wide open... Thank you for not poking me in the eye.

Chapter 1

Introduction

1.1 Objectives of research

The first objective of this research is to provide a comprehensive explanation of the theory behind the cell-transmission model for traffic flow in networks. The cell-transmission model was developed with contributions from diverse fields. This has led to gaps in the literature, and alterations of the model which contradict fundamental assumptions in the theory.

The second objective is to evaluate the accuracy of the cell-transmission model, and to investigate enhancements which may improve the accuracy. Of course, the implications of these enhancements on the simulation execution speed are considered, but it is of secondary importance in this work.

1.2 Approach methodology

The first objective requires a thorough investigation of the literature relating to the cell-transmission model. The cell-transmission model is quite recent, being defined as a solution method for the dynamic network loading problem for traffic networks by C.F. Daganzo in 1994. However, the cell-transmission model draws on early findings from many fields of study, such as traffic engineering (Greenshields, 1935), traffic flow theory (Lighthill and Whitham, 1955;

Richards 1956), numerical methods (Lax, 1954), and fluid dynamics (Godunov, 1959). Any gaps discovered in the literature were filled, and insights were included.

The cell-transmission model was evaluated qualitatively. The criteria were borrowed from Khattak and Jovanis (1990): the reliability of the theoretical foundations, the practicality of the methodology, the degree of policy sensitivity, the simplicity of the model, the data requirements, and the domain of application.

A quantitative evaluation of the cell-transmission model required an implementation, which is called CellNetLoad. A set of test networks were developed to determine the accuracy of the simulation results. Some small networks were used to isolate particular network features. A more complex test network was then used to evaluate the performance of the cell-transmission model on a realistic network. The evaluation includes an analysis of the impact of the discretisation of time-space on the simulation results. The simulation execution times were recorded for comparison.

The accuracy of CellNetLoad was determined by comparison with a benchmark. The benchmark was derived from a microscopic traffic simulation program, called INTEGRATION. Though the realism of INTEGRATION is questionable, it models traffic in greater detail than the cell-transmission model. So, it is reasonable to assume that it is closer to reality. Whenever possible this assumption is justified in the study with qualitative arguments based on quantitative analysis of the simulation results. There are several advantages of using a benchmark, which is derived from a microscopic simulation, instead of observations of an existing traffic network. The test network can be customised and altered to isolate certain features, the degree of stochasticity can be controlled, and the simulation is repeatable.

Based on the findings of the evaluation, several enhancements to the cell-transmission were implemented, then evaluated on the same test networks. These enhancements can be classified in two categories: enhancements to flow propagation, and enhancements to the intersection model. The enhancements to flow propagation include a non-linear equilibrium flow-density relationship, modelling arc strorage exactly, and the discretisation of flow. The enhancements to the intersection model involve the relaxation of the constraint that permits only diverge and merge intersections, the relaxation of first-in-first-out discipline, and the explicit modelling of lane striping at intersection approaches.

1.3 Structure of the thesis

As much as possible the document is structured such that each chapter is self-contained, though a thorough understanding of any chapter is enhanced by the knowledge and terminology gained by reading the preceding chapters. The terminology is summarised for ease of reference in the glossary at the front of the document, followed by a list of the variables used for the mathematical notation.

Chapter 2 contains a review of the theoretical foundation of the cell-transmission model. The contributions from various fields were integrated, and original insights were included. The qualitative evaluation of the cell-transmission model is at the end of chapter 2.

The CellNetLoad simulation model is described in chapter 3. The representation of the network and the data structures are described. Flow charts and algorithms were used to describe program flow. Although the enhancements to the cell-transmission model are mentioned, this chapter focuses on the use of CellNetLoad to simulate the basic cell-transmission model. The enhancements are described in chapters 5 and 6.

Chapter 4 includes the quantitative evaluation of the cell-transmission model. The simulation results of INTEGRATION and CellNetLoad, which include temporal flow, density, and travel time, are analysed and compared. Each of the test networks operates under uncongested and congested conditions. The networks include a linear roadway with a lane reduction, a diverge intersection, a merge intersection, and a 4-leg signalised intersection.

The enhancements to flow propagation are presented in chapter 5. The enhancements are described, and then the simulation results are repeated for the linear test network with the enhancements activated. The simulation results are analysed and compared to the results discussed in chapter 4. The impact of the enhancements on the simulation execution times is also included. An analysis of the computational complexity of the CellNetLoad model is performed. Regression analysis are used to determine the coefficients for the predicted simulation execution time for varying time-space discretisations. The predicted simulation execution time is compared to the simulation execution times with and without the flow-propagation enhancements activated.

Chapter 6 contains the description and evaluation of the enhancements to the generalised intersection model. The algorithms are presented for the relaxation of FIFO discipline, and for the explicit modelling of lane striping on intersection approaches. The convergence of the algorithm is proven and the upper bound on the number of iterations is found when possible. The simulation results are compared with those found in chapter 4 for the diverge intersection, the merge intersection, and the 4-leg intersection. The predicted simulation execution time is compared with the observed times for the simulations with and without the enhancements activated.

A more realistic test network is used to evaluate the cell-transmission model in Chapter 7. The network consists of a unidirectional freeway and a parallel arterial joined by an on-ramp and an off-ramp. The simulation results and the simulation execution times are compared with the benchmark and with the enhancements activated.

Chapter 2

Theoretical Foundation and Qualitative Evaluation of the Cell-transmission Model

2.1 Introduction

The cell-transmission model is used to model traffic flow on roadways. In recent years it has gained much popularity in research circles, partly because it is well suited to computer implementation. Its development was spurred by contributions from diverse fields, including fluid dynamics, thermodynamics, numerical methods, traffic flow theory, and traffic engineering. The main objective of this chapter is to review the theoretical foundations of the cell-transmission model by integrating the various contributions from specialists in each field using a consistent notation and terminology. (The notation is summarised in a separate section at the beginning of the document for easy reference.) Whenever they are useful, visual aids accompany the verbal explanations of traffic phenomena. It is hoped that the reader will gain an intuitive understanding of the complexities of traffic flow and how they are modelled.

The remainder of this chapter is structured as follows.

The second section describes the mathematical formulation of the dynamic network loading (DNL) problem, which is solved by the cell-transmission model. Alternative solution methods are categorised and described with reference to specific models.

The next section describes the development of the cell-transmission model. This section aims to define and describe the background theory which is essential for the comprehension of the formulation and critical discussion of the cell-transmission model.

The formulation of the cell-transmission model is given in the fourth section. The relationship between the cell-transmission model and the hydrodynamic model of traffic flow is explained. Some necessary adaptations to the cell-transmission model are introduced for the application to the DNL problem.

In the last section, a critical discussion of the cell transmission model draws comparisons with alternative solution methods for the DNL problem.

2.2 Approaches to Dynamic Network Loading

A network $\tilde{G} =< N, A >$ is represented as a graph $G =< N, A >$ where N is the set of nodes and A is the set of arcs. In addition to the graph topology, the network also contains the node and arc attributes. Arcs are defined as $a = (n_a^-, n_a^+)$, such that $n_a^-, n_a^+ \in N$ are the head (upstream node) and tail (downstream node) of arc a, respectively. The traffic demand $g_r(t)$ is expressed as a flow rate at departure time t on route $r = \{a_1, a_2, \ldots, a_r\} \in R_p$, where $\|r\|$ is the number of arcs on route r; $a_1, a_2, \ldots, a_r \in A$ are the individual arcs, such that $n_{a_1}^- \in O$ and $n_{a_r}^+ \in D$; R_p is the set of routes for origin-destination pair p, $p \in O \times D$; and $O \subset N$ and $D \subset N$ are the sets of origin and destination nodes, respectively.

The goal of dynamic network loading (DNL) is the determination of the set of temporal arc flows $\{q_a(t)\}\ \forall a \in A\ \forall t$ and the set of temporal arc travel times $T(t) = \{T_a(t)\} \forall a \in A\ \forall t$ for a given network $\tilde{G}$ and known temporal path flows $g_r(t)\ \forall r \in R_p \forall p\ \forall t$. Travel times and path flows correspond to the arc or path entrance time. The time period under consideration is $[0,\bar{T}]$, where the end of the interval is given The dependence of arc travel times on the network and on path flows can be expressed as $T_a(t) = f(\tilde{G}, g_r(z)) \forall r \in R_p\ \forall p\ \forall z \in [0,t)$. The arc travel times can be used to estimate any path travel time by summing the travel times of the constituent arcs.

In a broader context than DNL, the path flows and arc travel times are interdependent. The set of temporal path flows is an elastic quantity which may change due to the resulting path travel times, i.e. $g_r(t) = f(T(z)) \forall z \in [0,\bar{T}]$. The temporal path flows may be affected by the arc travel times in several ways. For example, in order to achieve an equilibrium assignment, flow may be shifted from longer to shorter paths or to previously unused paths which connect the same origin-destination pair. According to a behavioural model, some flow may have the destination changed to one that is closer to the departure origin, or some trips may be cancelled.

Thus, the proper determination of both arc travel times $T(t) \forall t \in [0,\bar{T}]$ and temporal path flows $g_r(t) \forall t \in [0,\bar{T}]$ requires (1) a traffic demand model which predicts temporal path demand flows from given travel times, (2) a DNL model which predicts travel times from given temporal path demand flows, (3) an initial set of temporal path demand flows, and (4) an iterative solution algorithm which converges to a feasible solution of the stated model, which may be called the equilibrium state. The dependence of temporal path flows on travel times and the iterative solution of dynamic traffic assignment models are beyond the scope of this work. The DNL model is the focus. Temporal path demand flows are assumed to be given.

Many models and solution strategies have been considered to solve the problem of dynamic network loading, each with a different set of simplifying assumptions and its associated benefits and drawbacks. The classification of solution methods is a contentious endeavour. The difficulty in gaining acceptance of some sort of classification scheme is rooted in the variety of objectives that are sought from different perspectives. One party may value precision of travel time estimates as the most important criteria, while another party seeks computational efficiency on large networks. While practitioners may be concerned with issues such as data requirements and calibration, academic researchers may be more appreciative of mathematical elegance. Some applications require only the simplest of arc and node models, while others necessitate more realistic models for certain network features (such as lanes reserved for high occupancy vehicles, toll roads, and features of intelligent transportation systems), or models which account for particular traffic phenomena (such as queue spillback, lane changing, mixed vehicle fleets, and vehicle acceleration/deceleration). The wide range of choices explains the existence of the vastly different models and solution strategies to the dynamic network loading problem.

Why bother to classify the various models and solution strategies? The task of gaining an understanding of the different approaches requires some grouping of similar models. Furthermore, comparison of the effects resulting from various modelling assumptions is greatly aided by some structure. It is also beneficial to have a classification scheme to determine which models are best suited for a particular domain of study or application. For these reasons, four categories are proposed, in order of increasingly detailed models: analytical models, macroscopic models, mesoscopic models, and microscopic models.

The borders between each category are fuzzy at best. Nevertheless, in the remainder of this section divisions have been made, and each category of DNL model is characterised based on the representation of the traffic stream, the treatment of space and time, and the level of detail in traffic dynamics which is modelled.

2.2.1 Analytical dynamic network loading models

Analytical DNL models use whole links as the space discretisation unit of the network. The traffic stream is represented by a piecewise continuous flow at the entrance and exit of each link. The link outflow is a function of the link inflow and/or link occupancy. The link outflow is given by either a deterministic exit link function or a deterministic travel time function. Conservation laws are respected on links and flows are balanced at nodes.

Codina and Barceló (1995) surveyed the development of analytical dynamic network loading models within the context of the dynamic traffic assignment problem. These analytical models have evolved from generalisations of the static equilibrium model for the traffic assignment problem (Florian and Hearn, 1995). The following discussion applies only to the features of each model which affect DNL. Therefore, issues relevant to finding equilibrium solutions to dynamic traffic assignment are ignored.

Merchant and Nemhauser's (1978a, 1978b) model was developed for networks with a single destination. The time interval under consideration is discretised into subintervals. The link exit flows during each time subinterval are given by exit link functions which depend solely on the link occupancy at the beginning of the subinterval. These exit link functions are continuous, non-decreasing, concave functions that reach the link outflow capacity as the link occupancy tends to infinity. Extensions to multiple destination networks violated FIFO discipline, which means that vehicles which enter a particular link heading to a particular destination may exit that link before vehicles that are already on that link heading to a different destination. In reality, link traversal times generally should be approximately equal for all vehicles which enter the link during the same short period of time, independent of their respective destinations. Carey (1991) formulated constraints for this model which ensure that FIFO discipline is respected for networks with multiple destinations, but at the expense of greatly complicating the solution algorithm.

Several continuous-time DNL models have been proposed. Friesz et al. (1989) generalised the Merchant and Nemhauser model into a continuous-time formulation using optimal control theory with linear exit functions, which imply infinite flow for infinite link density. Wie et al. (1990) extended this model to networks with multiple destinations. Boyce et al. (1992) and Ran et al. (1991) also developed continuous-time optimal control models, but without exit link functions. Though these models respect FIFO discipline, they imply the instantaneous propagation of flows along links. Xu et al. (1999) avoided instantaneous link traversal times by formulating a continuous-time DNL model which replaces the exit link functions with link traversal time functions, which are positive, non-decreasing, continuously differentiable functions of the link occupancy only at the instant when a vehicle enters the link. The model includes a FIFO constraint. Path departure rate functions must be non-negative, finite, and Lebesgue integrable. The unique solution of the DNL problem can be found in a finite number of iterations by the event-based DYNALOAD solution algorithm.

Another DNL model was formulated as a variational inequality by Drissi-Kaïtouni (1990) and Drissi-Kaïtouni and Hameda-Benchekroun (1992) within the context of the dynamic traffic assignment problem for networks with multiple destinations. The instantaneous link traversal time associated with link exit functions is avoided using a travel time function, which consists of two components: the fixed duration, which represents the free-flow travel time; and the delay due to congestion, which takes place in a vertical queue prior to discharge from the link. Congestion constraints are included for link inflow, link outflow, and link storage capacity, but FIFO discipline is violated if the solution has active congestion constraints (Codina and Barceló, 1995).

Astarita (1996) and Astarita et al. (1998a) improved the use of travel time functions in a continuous-time model for networks with multiple destinations. FIFO discipline is always respected. Each link consists of two segments: a running segment and a queuing segment. Packets of vehicles travel at the free-speed on the running segment until they reach the queuing segment which is a horizontal queue at jam density. Each link has inflow, outflow, and storage capacities. An important feature of this model is that the queuing segment can accommodate congestion which spills back from downstream links. Several node models were designed by Astarita et al. (1998b) which employ constants to partition downstream local supply to the upstream local demand. An event-based algorithm is used to solve the model.

2.2.2 Macroscopic dynamic network loading models

In macroscopic DNL models, vehicles are modelled with piecewise continuous functions of density and flow in space-time, analogous to fluid flow. The mathematical theory behind these models is based on temporal, one-dimensional fluid dynamics. Conservation laws are respected. The state variable(s), density (and sometimes speed), is described over the entire length of each link. Flow and speed are functions of the state variable(s). Also, macroscopic models assume homogeneous traffic characteristics and deterministic modelling relationships. Flow and density are related by empirical measurements (Greenshields, 1935).

The simplest macroscopic traffic models are based on the concept that vehicle traffic at a given point in space-time is affected only by conditions within a neighbourhood of that point (Daganzo, 1995b). Therefore, the future state at one instant later at a given point in space-time depends only on conditions at that point and the conditions immediately upstream and downstream of that point.

These models are described mathematically by a system of first order partial differential equations (PDEs). The LWR (Lighthill and Whitham, 1955; Richards, 1956) hydrodynamic model is a well-known first order PDE. Other first order systems include Bick and Newell (1960) for two-lane bi-directional roads and Munjal and Pipes (1971) for multilane freeways. Density is the single state variable. All first order PDE macroscopic models assume that traffic is always in an equilibrium state, which means that the flow is always the same for any given density.

A higher order PDE model was developed by Payne (1971) which explicitly models reaction time and anticipation of downstream traffic conditions. The state variables are density and speed. The traffic stream may be in a dynamic state, i.e. the flow may be different from the equilibrium flow for a given density. Payne's (1971) model converges to the LWR hydrodynamic model as the reaction time tends to zero, which implies instantaneous acceleration and deceleration. Several other higher order PDE models were based on kinetic theory of vehicular traffic developed by Prigogine and Hermann (1971), namely Phillips (1979), Helbing (1997), and Kerner et al. (1996).

Though PDE models can often be solved exactly using analytical techniques, complications due to multiple solutions for certain realistic initial value problems make these analytical approaches unwieldy. Consequently, the solution process has been mechanised for computation using numerical methods, namely approximations using finite difference equations (FDEs). The cell-transmission model is an FDE, which approximates the LWR hydrodynamic model. Other FDEs which approximate PDEs include Payne (1979), Michalopoulos et al. (1984a, 1984b), and Cremer and May (1985).

2.2.3 Microscopic dynamic network loading models

Microscopic DNL models represent vehicles discretely, often with varying characteristics and/or multiple classes. Space is continuous, and time is finely discretised. Vehicle positions are updated using car-following logic (Gazis, 1974; Gabard, 1991) and lane-changing rules, which are often stochastic. Variability in driver behaviour and/or vehicle dynamics is explicitly modelled. Interactions between vehicles at intersections may be modelled by right-of-way and gap-acceptance rules. These models often employ stochastic modelling relationships and heuristic rules.

Microscopic traffic simulation models have been developed by the traffic engineering community. Though early attempts required mainframe computers, today they are usually designed for use on a personal computer. TRAF-NETSIM (Yedlin et al., 1994), sponsored by the Federal Highway Administration, is the result of merging two distinct models for freeway networks and urban networks. INTEGRATION (Van Aerde, 1999a) which began as a mesoscopic model, evolved into a microscopic traffic simulator based on a speed-headway relationship that is derived from macroscopic flow parameters (Van Aerde, 1995). INTEGRATION models all types of roadways using consistent logic that is generalised to accommodate freeways and surface streets.

Other commercial models are AIMSUN2 (Ferrer and Barceló, 1993; Barceló et al., 1998), and VISSIM (Fellendorf, 1994). Some models, which are primarily used for research or particular applications include MITSIM (Yang and Koutsopoulos, 1996), NEMIS (Mauro, 1991), DRACULA (Liu and van Vliet, 1996), and SITRA-B+ (Omli and Farges, 1992).

2.2.4 Mesoscopic dynamic network loading models

Mesoscopic DNL models have a variety of forms and modelling assumptions. They can be loosely defined as all models that are more detailed than macroscopic models and more computationally efficient than microscopic models. In general, vehicles are modelled discretely or grouped into packets, which may have heterogeneous characteristics. Space is often continuous. Time may be continuous, or discretised coarsely or finely. Both deterministic and stochastic modelling relationships are common.

In DYNAMIT (Ben-Akiva and Bierlaire, 1999) time is discretised and each link is represented by one segment, or a series of segments, each with a moving part and a queuing part. Vehicles with homogeneous characteristics are aggregated into indivisible packets, which may overlap one another on a segment. A packet enters a moving segment at a speed derived from the speed-density relationship. Then the packet undergoes constant acceleration or deceleration until it reaches the end of the segment at a speed which is consistent with the entrance speed to the next segment. But, if there is a queue on a queuing segment, then the packet must decelerate to the discharge speed from the queue. The inflow and outflow capacity of each segment must be respected. The capacities, densities and speeds are updated for each segment after each update phase. Several advance phases occur within each update phase, during which the packets move according to the above-mentioned rules.

DYNASMART (Jayakrishnan et al., 1994) also discretises time and models links with one or more segments, each with a moving part and a queuing part. Vehicles are usually modelled discretely though grouping into homogeneous packets called macroparticles (Chang et al., 1985) is permitted. Vehicle speed on a segment is a non-linear, deterministic function of density with free-speed at zero density and positive minimum speed at jam density. The position of each vehicle is decreased according to the distance travelled during a time-step until the vehicle reaches the back of the horizontal queue, which is stored at jam density at the end of the segment. On surface streets, a node model determines the exit volume from the queue, whereas vehicles automatically advance to the next freeway segment. The node model respects exit, entrance, and storage capacities of the segment, which account for the signalisation explicitly as well as the effects of gap-acceptance for left turns. After a vehicle is released from the queue, the position on the downstream segment is updated according to the remainder of the time-step which is available for travel at the speed of the downstream segment. The travel time estimate is based on the current speed and non-queued length of the moving part and the queue size and a moving average of the service rate of the queuing part of each segment. Vehicle positions are not required to respect the jam density on the moving part of the segment. No lane-changing, car-following, or platoon dispersion is modelled. FIFO discipline is respected.

Another mesoscopic DNL model is based on particle hopping rules (Nagel, 1998). Vehicles are represented discretely at discrete positions and times, which implies discrete vehicle speeds, since the speed is the change in position divided by the change in time. Deterministic versions of these models have analogous macroscopic models in the limit of finely discretised vehicles, positions, and times. Stochastic particle hopping rules account for variability in driver behaviour and vehicle characteristics, but the analogy to macroscopic traffic flow models is lost. Cellular automata models are particle hopping models in which the hopping rules are myopic, such that each cell could be updated in parallel. TRANSIMS (Nagel et al., 1998) includes a cellular automata model of traffic flow in networks with stochastic particle hopping rules.

Another class of mesoscopic model is based on queuing theory. These models differ from the analytical models in two ways: (1) vehicles are discrete, and (2) subtle features of traffic flow are modelled, such as lane-changing on links and prioritised movements at controlled intersections. One such model is currently under development (Mahut, 1999).

A queuing-based mesoscopic sub-model has been developed recently within INTEGRATION for the purpose of improving assignment prior to the microscopic simulation (Van Aerde, 1999b). Vehicles are still represented discretely in this sub-model, but queuing logic is substituted for the car-following logic of the microscopic model.

2.3 Development of the Cell-transmission Model

The theoretical groundwork of the cell-transmission model has been developed from several different perspectives. In addition to providing the name for the model, Daganzo (1994, 1995a, 1995b) described the efficiency and practicality of the cell-transmission model from a traffic engineering perspective. The cell-transmission model was presented as an approximate solution to the LWR hydrodynamic model (Daganzo, 1995b). Lebacque (1996) showed that the cell-transmission model fits into a broader context as a solution to an initial value problem of a hyperbolic conservation equation which is similar to the Godunov (1961) method.

This section presents the LWR hydrodynamic model, an analytical solution method, and some finite difference approximations. The measurement of speed-flow-density parameters and their use in flow-density relationships, and the derivation of the conservation law lead to the LWR hydrodynamic model. An understanding of kinematic waves and shockwaves is essential to grasp the analytical solution method, called the method of characteristics. The last section introduces concepts of finite difference approximations to the LWR hydrodynamic model, which served as the foundation for the cell-transmission model.

2.3.1 Speed-flow-density relationship

A long stretch of road without merges, diverges, or intersections can be characterised by a flow-density relationship that is quantified empirically (Lighthill and Whitham, 1955). The flow q at time t on the slice of the road of length d_x, centred at x, is approximated by

[2.1] $$q(x,t) \cong \frac{n}{\Delta t}$$

where n is the number of vehicles crossing the slice of road in time interval Δt, such that t is the middle of interval Δt. The density k (also know as concentration) at time t and location x is the average number of vehicles on the slice of road divided by its length

[2.2] $$k(x,t) \cong \frac{\sum_n d_t(n) / \Delta t}{d_x}$$

where $d_t(n)$ is the time taken by vehicle n to cross the slice of road of length d_x. The space-mean speed v is the average vehicle speeds weighted according to the time for each vehicle to cross the slice of road, which is equivalent to the ratio of flow to density

[2.3] $$v = \frac{nd_x}{\sum_n d_t(n)}$$

(The space-mean speed is different from the time-mean speed, which is the unweighted average speed of the n vehicles.)

Substitution of [2.1] and [2.2] into [2.3] is consistent with the fundamental speed-flow-density relationship

[2.4] $$q = vk,$$

which can be used to obtain the value of the third parameter, given the values of the other two parameters.

2.3.2 Equilibrium flow-density curves

The relationships between flow and density, or speed and density, are often used to model traffic streams. The specification of either relationship implicitly determines the other relationship using [2.4]. Speed-density relationships are often used for microscopic DNL models since they are closely related to speed-headway relationships and car-following models. Flow-density relationships are more commonly applied in macroscopic DNL models since they are useful, in association with space-time diagrams, for analysis of kinematic waves and shockwaves, which will be discussed later.

Greenshields (1935) was the first to document a relationship between speed and headway as a straight line, based on empirical evidence. The maximum speed, called the free speed v_f, occurs at zero density. The maximum density, called the jam density k_{jam}, occurs at zero speed. With appropriate substitutions from [2.4], this linear speed-density relationship is equivalent to parabolic flow-density and speed-flow relationships (Figure 2.1). In fact, any relationship between two of the state variables implies corresponding relationships between the other two pairs of state variables. The maximum flow $q_{\max}$ occurs at the critical speed $v_c = v_f / 2$ and critical density $k_c = k_{jam} / 2$, which implies that $q_{\max} = v_f k_{jam} / 4$. The model can be described by two parameters: the free speed v_f and the jam density k_{jam}.

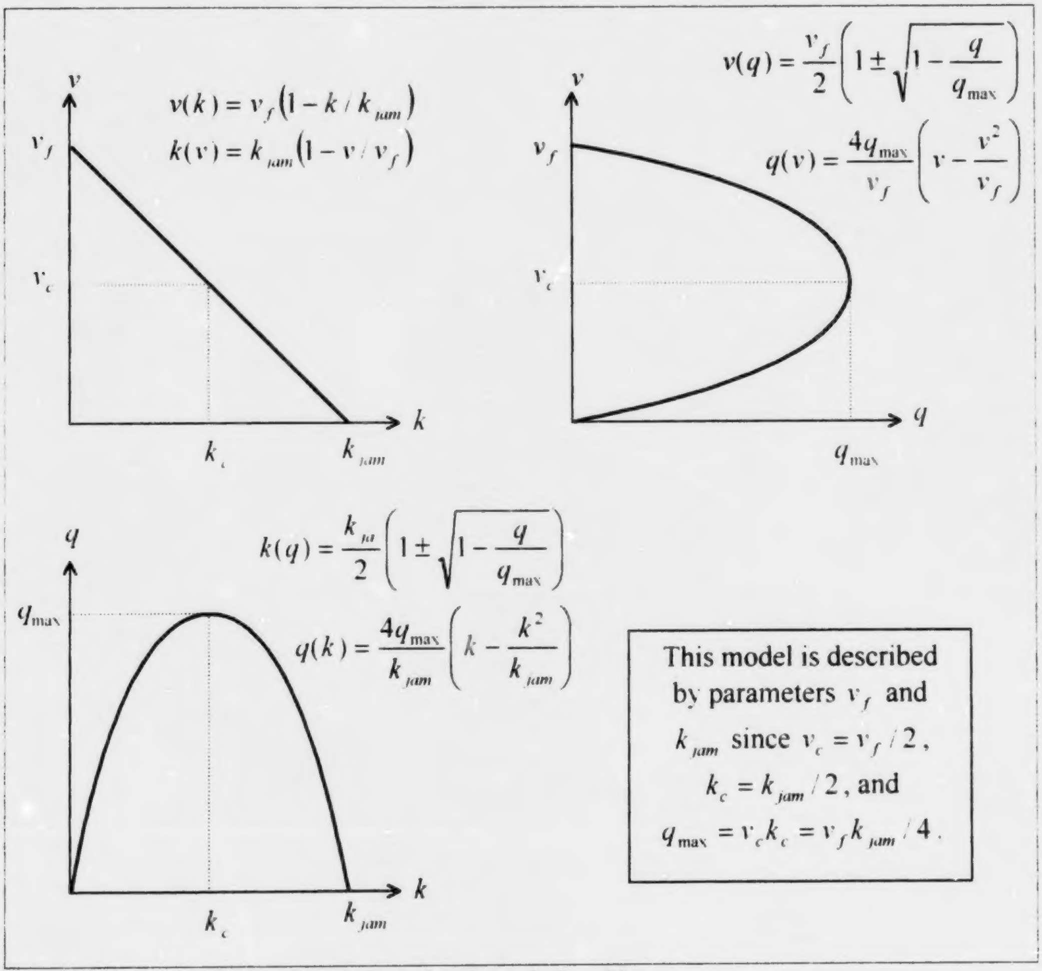


Figure 2.1 Greenshields speed-flow-density relationship

This model is used to characterise traffic states under steady-state conditions only. Given the density, the model can produce the equilibrium flow and equilibrium speed. Likewise, given the speed, the model can produce the equilibrium flow and equilibrium density. Given only the flow, however, the model can produce two possible equilibrium flow and equilibrium density values.

The equilibrium flow-density relationship warrants the closest attention for the purposes of macroscopic dynamic network loading models. Relaxation of Greenshields' assumption that speed and density are linearly related yields the generalised equilibrium speed-density relationship. The corresponding generalised equilibrium flow-density curve then loses the parabolic shape (Figure 2.2). Four parameters are used to describe the flow-density relationships:

free speed v_f,

jam density k_{jam},

maximum flow q_{max}, and

critical density $k_c = q_{max} / v_c$, where v_c is the critical speed.

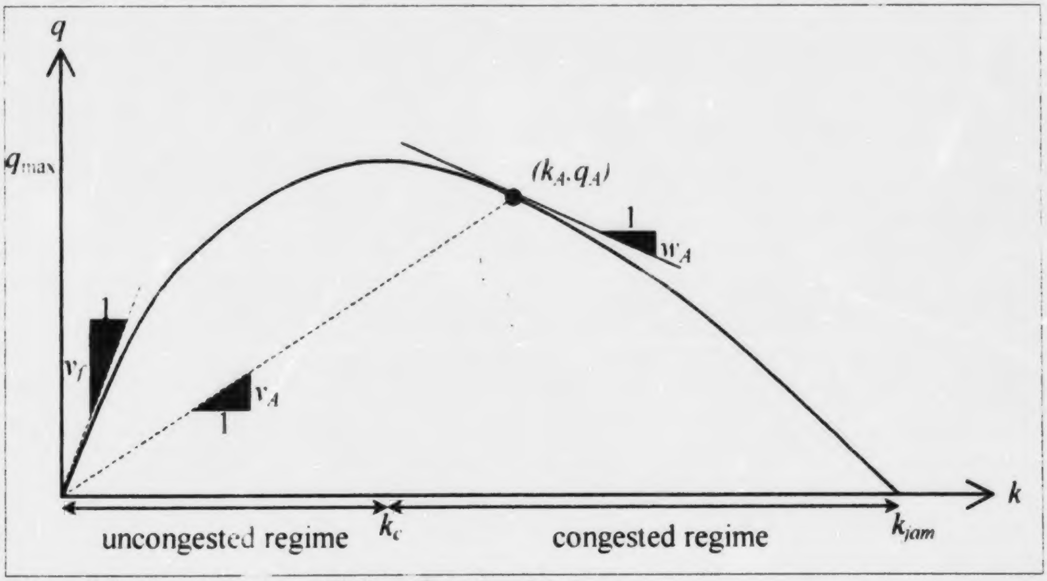


Figure 2.2 Generalised equilibrium flow-density curve

Note from [2.4] that the vehicle speed is equal to the slope of a radius vector from the origin to the relevant point on the equilibrium flow-density curve, as shown on Figure 2.2. Therefore, the free-speed is the slope of the tangent to the speed-density curve at zero density.

The equilibrium flow-density relationship can be described by two regimes. As the density approaches zero, the flow must also approach zero. As the density approaches the maximum density, called the jam density k_{jam}, drivers decrease their speed to ensure a safe stopping distance, thereby causing the flow to approach zero. The portion of the curve on which flow *increases* with density is called the *uncongested regime* of the equilibrium flow-density curve. The portion of the curve on which flow *decreases* with density is called the *congested regime*. These two regimes are separated by the maximum flow $q_{\max}$ at the critical density k_c.

The flow-density curve for a stretch of road may change with position x due to roadway geometry, such as narrowing of the roadway at a bridge, and with time t due to lane blockages, which often result from incidents or construction. In general, the equilibrium flow-density relationship is $\tilde{q}(k,x,t)$. This generalised relationship has been specified using a variety of functional forms, each with a set of parameters.

2.3.2.1 Non-linear equilibrium flow-density relationships

A wide variety of functions have been proposed to characterise the equilibrium flow-density relationship (May, 1990). Since it is impossible to observe purely steady-state conditions, empirical evidence cannot be used to determine the true equilibrium curve. Nevertheless, field measurements are used to justify the appropriateness of various mathematical functions to describe the equilibrium flow as a function of density.

The Greenshields (1935) parabolic model was the first of many continuous function equilibrium flow-density functions. Greenberg (1959) found that $\tilde{q}(k) = v_c k \ln(k_{jam} / k)$, which has infinite free speed. Underwood (1961) proposed $\tilde{q}(k) = v_f k \exp(-k / k_c)$, which has infinite jam density. Drake et al. (1967) suggested a similar model $\tilde{q}(k) = v_f k \exp(-(k / k_c)^2 / 2)$. Van Aerde (1995) proposed a more complex model which uses all four parameters v_f, k_{jam}, $q_{\max}$, and v_c. These models are depicted in Figure 2.3 for a consistent set of values for parameters v_f, k_{jam}, $q_{\max}$, and k_c, with $v_c = q_{\max} / k_c$. Each of these models uses a continuous mathematical function to describe the equilibrium flow in both the congested and uncongested regimes.

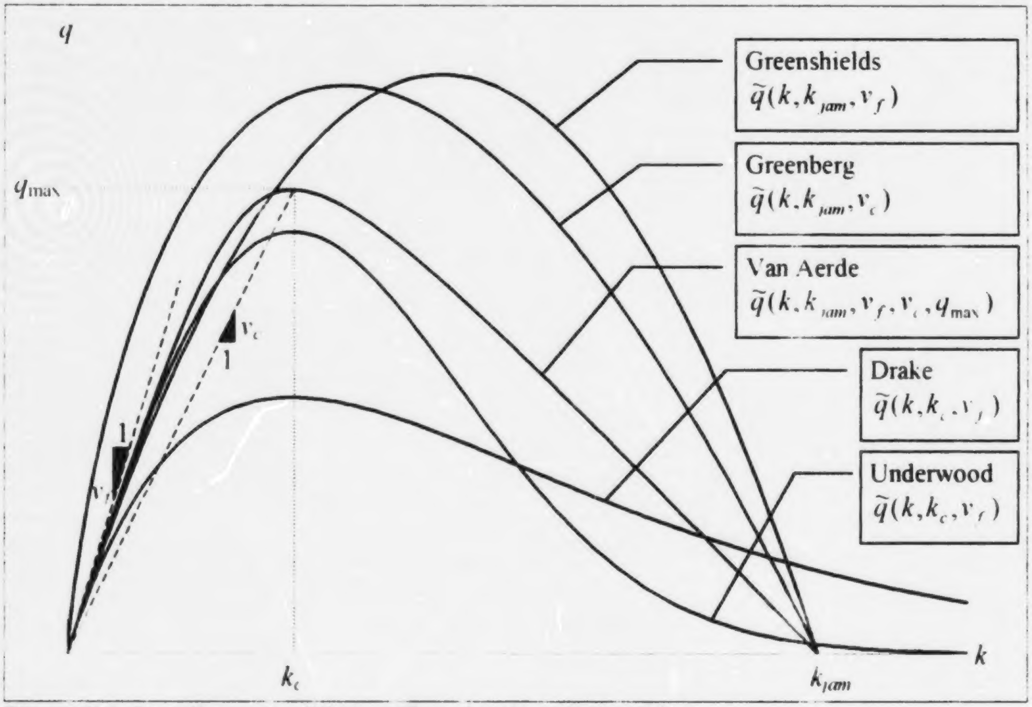


Figure 2.3 Non-linear equilibrium flow-density relationships

In order to maintain relatively simple mathematical functions while improving the fit of the equilibrium flow-density curve to observed data, several multi-regime models have been proposed with a different function in each density regime. Edie (1961) chose the Underwood model for the uncongested regime, and the Greenberg model for the congested regime. Drake et al. (1967) suggested the use of separate Greenshields models for each of three regimes: the uncongested regime, the congested regime, and a transitional regime. The boundaries between the regimes were identified using maximum likelihood functions with data from observations of traffic flow.

2.3.2.2 Linear equilibrium flow-density relationships

Linear equilibrium flow-density relationships are generally considered to be approximations of non-linear relationships. They could also be considered as extensions of the principle which yielded the multi-regime non-linear models: simpler mathematical functions on multiple density regimes. The linearity is exploited to simplify mathematical formulation of some traffic flow models. Each of the following linear relationships is depicted in Figure 2.4.

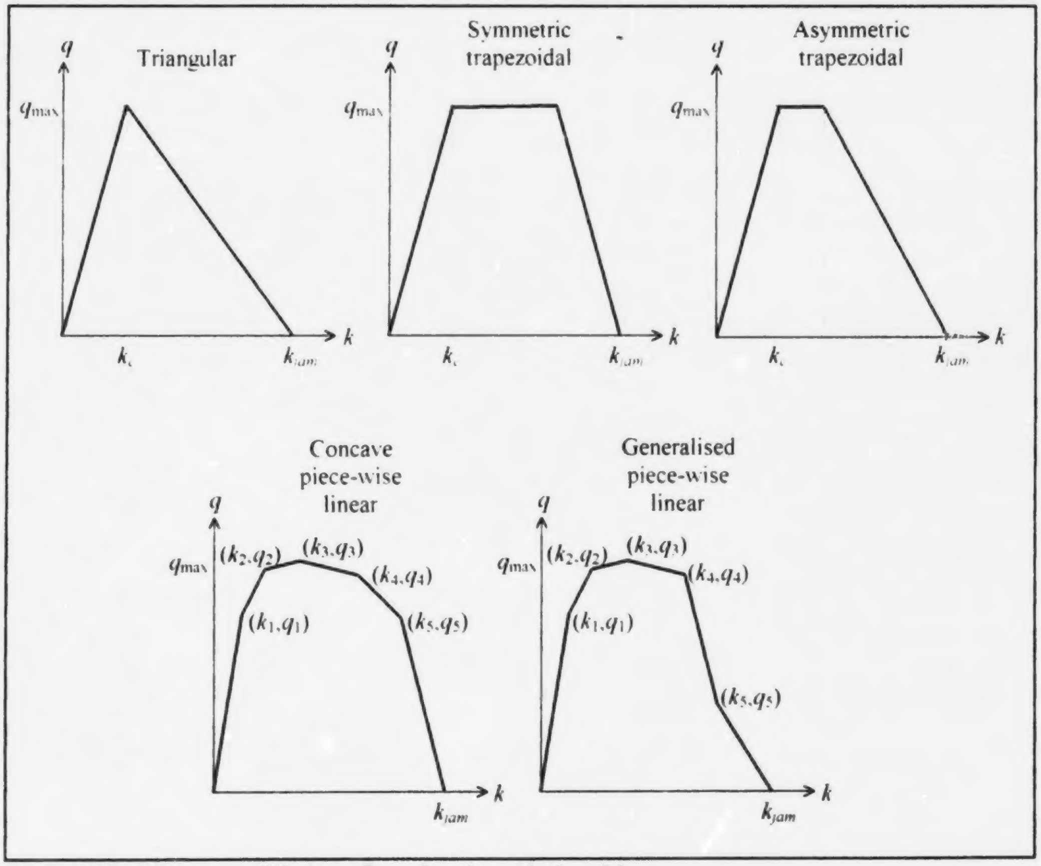

Triangular
Symmetric trapezoidal
Asymmetric trapezoidal
q
q_max
k
k_c
k_jam
Concave piece-wise linear
Generalised piece-wise linear
(k_1,q_1)
(k_2,q_2)
(k_3,q_3)
(k_4,q_4)
(k_5,q_5)

Figure 2.4 Linear equilibrium flow-density relationships

Newell (1993) proposed a simplified triangular flow-density relationship

$$\tilde{q} = \min\left(k\frac{q_{\max}}{k_c}, (k_{jam} - k)\frac{q_{\max}}{k_{jam} - k_c} \right)$$

with parameters jam density k_{jam}, critical density v_c, and maximum flow $q_{\max}$. Note that the free speed $v_f = q_{\max} / k_c$ is the equilibrium speed throughout the uncongested regime.

Daganzo (1994) described a symmetric and an asymmetric trapezoidal flow-density relationship. The symmetric trapezoid is essentially a truncated isosceles triangular relationship

$$\tilde{q} = \min\left(k\frac{q_{\max}}{k_c}, q_{\max}, (k_{jam} - k)\frac{q_{\max}}{k_c} \right).$$

In the asymmetric trapezoidal flow-density relationship

$$\tilde{q} = \min\left(k\frac{q_{\max}}{k_A}, q_{\max}, (k_{jam} - k)\frac{q_{\max}}{k_{jam} - k_B} \right)$$

the slope in the congested portion is specified as $q_{\max} /(k_B - k_{jam})$, using an additional critical density k_B to the previously defined critical density $k_c = k_A$.

A piecewise-linear approximation of the equilibrium flow-density curve allows for greater flexibility than the above models, though it requires the specification of a series of data points (k, q), the first of which must be $(0,0)$, and the last being $(k_{jam}, 0)$. This piecewise-linear equilibrium flow-density relationship is usually a non-convex model. However, Daganzo (1999) employs a concave portion in the congested regime.

The flow-density relationship is combined with the conservation law of fluid dynamics to form the LWR hydrodynamic model. The conservation law is derived in the next section. The subsequent section describes the LWR hydrodynamic model.

2.3.3 Conservation law

The conservation law of fluid dynamics asserts that the rate of change of the total amount of substance contained in a fixed domain is equivalent to the flux (directional flow) of that substance across the boundary of that domain (Lax, 1973).

In the context of road traffic, the rate of change during the time interval $[t_1, t_1 + \Delta t]$ of the number of vehicles on a domain, defined as the section of roadway $[x_1, x_1 + \Delta x]$, can be described as

$$\int_{x_1}^{x_1+\Delta x} \left[k(x,t_1+\Delta t) - k(x,t_1)\right]\mathrm{d}x = \int_{t_1}^{t_1+\Delta t} \left[q(x_1,t) - q(x_1+\Delta x,t)\right]\mathrm{d}t .$$

Now, division of both sides by $\Delta x \Delta t$ and taking limits as Δx and Δt approach zero leads to

$$\lim_{\Delta t\to 0}\left(\lim_{\Delta x\to 0}\frac{\int_{x_1}^{x_1+\Delta x}\left[k(x,t_1+\Delta t)-k(x,t_1)\right]\mathrm{d}x}{\Delta x\Delta t}\right) = \lim_{\Delta x\to 0}\left(\lim_{\Delta t\to 0}\frac{\int_{t_1}^{t_1+\Delta t}\left[q(x_1,t)-q(x_1+\Delta x,t)\right]\mathrm{d}t}{\Delta x\Delta t}\right)$$
$$\Rightarrow \lim_{\Delta t\to 0}\frac{k(x_1,t_1+\Delta t)-k(x_1,t_1)}{\Delta t} = \lim_{\Delta x\to 0}\frac{q(x_1,t_1)-q(x_1+\Delta x,t_1)}{\Delta x}$$
$$\Rightarrow \frac{\partial k(x_1,t_1)}{\partial t} = -\frac{\partial q(x_1,t_1)}{\partial x}.$$

Similarly, this finding can be replicated for any point in space-time (x,t). Therefore, the expression of the conservation law for vehicles in a traffic stream is

[2.5] $$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 .$$

Note that the conservation of vehicles is independent of any flow-density relationship.

2.3.4 Lighthill-Whitham-Richards hydrodynamic model

To apply the analogy of fluid dynamics to road traffic, discrete vehicles on a roadway may be approximated by a continuous density $k(x,t)$ of vehicles, which varies with position x and time t. This density defines the state of the system in space-time. The flow $q(x,t)$ at each point in time-space on the roadway depends on an equilibrium flow-density relationship given by

[2.6] $$q(x,t) = \tilde{q}(k(x,t),x,t) .$$

Substitution of [2.6] into the conservation law [2.5] yields

[2.7] $$\frac{\partial k(x,t)}{\partial t} + \frac{\partial \tilde{q}(k(x,t),x,t)}{\partial x} + \frac{\partial \tilde{q}(k(x,t),x,t)}{\partial k(x,t)} \frac{\partial k(x,t)}{\partial x} = 0 .$$

This situation corresponds to a fluid flowing in a non-uniform—changing its cross-sectional size and/or shape along its length with time—pipe, hence the name "hydrodynamic model."

It can be observed from [2.7] that the hydrodynamic model is a first order model. All flow states must lie directly on the equilibrium flow-density relationship specified in [2.6]. Any change in flow state must either follow the equilibrium flow-density curve, or jump to another flow state on the curve. Such a jump implies that vehicles which traverse a density discontinuity in the space-time plane immediately take on the equilibrium flow associated with the density in the new flow state, which implies an instantaneous acceleration or deceleration according to [2.4]. Therefore, the hydrodynamic model assumes that acceleration and deceleration rates are unbounded.

Lighthill and Whitham (1955) showed how [2.7] is related to kinematic waves and shockwaves, which are discussed in the following sections. They also analysed in detail the case in which the equilibrium flow-density relationship was stable with respect to time t. Given initial conditions $k(x,0)\ \forall x$ Richards (1956) solved [2.7] analytically for a homogeneous roadway with a parabolic equilibrium flow-density curve. The names Lighthill-Whitham-Richards model, hydrodynamic model, and kinematic wave model have become synonymous.

2.3.5 Kinematic waves

A kinematic wave is defined as a continuous series of points in space-time along which the flow is constant

[2.8] $$\frac{\mathrm{d}q(x,t)}{\mathrm{d}t} = 0$$

By expanding [2.8] into partial derivatives, it can be shown that the directional derivative of the flow in the space-time plane is zero along the path of the kinematic wave

$$\begin{aligned} &\frac{\partial q(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} = 0 \\ &\Rightarrow \frac{\partial q(x,t)}{\partial k(x,t)}\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} = 0 \end{aligned}$$

Substituting $\frac{\partial q(x,t)}{\partial x} = -\frac{\partial k(x,t)}{\partial t}$ from the conservation of vehicles [2.5], yields a relation between the trajectory in the space-time plane and the flow-density relationship.

$$\begin{aligned} &\frac{\partial q(x,t)}{\partial k(x,t)}\frac{\partial k(x,t)}{\partial t} - \frac{\partial k(x,t)}{\partial t}\frac{\mathrm{d}x}{\mathrm{dt}} = 0 \\ &\Rightarrow \frac{\mathrm{d}x}{\mathrm{dt}} = \frac{\partial q(x,t)}{\partial k(x,t)} \end{aligned}$$

The kinematic wave speed w is the change in position with time along the kinematic wave, hence

[2.9]
$$w(x,t)=\frac{\partial q(x,t)}{\partial k(x,t)}.$$

If it is assumed that flow is related to density according to the equilibrium flow-density curve, then [2.9] can be specified as

[2.10]
$$w(x,t)=\frac{\partial \tilde{q}(k(x,t),x,t)}{\partial k(x,t)}.$$

Graphically, the kinematic wave speed w is equal to the slope of the tangent of the equilibrium flow-density curve at position x and t, as shown in Figure 2.2. So in the uncongested regime ($k < k_c$) the wave speed is always non-negative; in the congested regime ($k > k_c$) the wave speed is always non-positive; and at the critical density ($k = k_c$) the wave speed must be zero.

Also, the wave speed w can be compared to the vehicle speed v by substituting the fundamental speed-flow-density relationship [2.4] into [2.9], resulting in

[2.11]
$$w(x,t)=\frac{\partial\left(v(x,t)k(x,t)\right)}{\partial k(x,t)}=v(x,t)+k(x,t)\frac{\partial v(x,t)}{\partial k(x,t)}$$

Since speed is generally found to decrease with density, a consequence of [2.11] is that the kinematic wave speed w may not exceed the speed of the vehicles in the traffic stream v

$$\frac{\partial v(x,t)}{\partial k(x,t)}\leq 0 \Rightarrow w \leq v.$$

The variation of the density along the path of the kinematic wave is obtained by substitution of

$\frac{\mathrm{d}x}{\mathrm{d}t} = w(x,t) = \frac{\partial \tilde{q}(k(x,t),x,t)}{\partial k(x,t)}$ from [2.10] into [2.7], as follows

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial \tilde{q}(k(x,t),x,t)}{\partial x} + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial k(x,t)}{\partial x} = 0 .$$

and rearrangement yields the directional derivative of the density in the space-time plane along the path of the kinematic wave

[2.12] $$\frac{\mathrm{d}k(x,t)}{\mathrm{d}t} = \frac{\partial k(x,t)}{\partial t} + \frac{\partial k(x,t)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\partial \tilde{q}(k(x,t),x,t)}{\partial x}$$

So, the rate of change in density along the path of a kinematic wave is equivalent to the rate of change in the equilibrium flow along the roadway.

The kinematic wave can be visualised as the trajectory traced by a motorcycle driving at speed $w(x,t) = \frac{\partial \tilde{q}(x,t)}{\partial k(x,t)}$ on the shoulder of a road at position x and time t, such that it does not interfere with the traffic stream on the road. The flow remains constant at any point along the path of the motorcycle, so the trajectory of the motorcycle on a distance-time graph is an isoflow line, also known as a characteristic. Note that the characteristic (and the motorcycle) is permitted to travel with or against the direction of flow, or it may remain stationary.

Different cases of the way the equilibrium flow-density relationship varies with position and time yield special properties of kinematic waves. If the equilibrium flow-density curve $\tilde{q}(k(x,t),x,t)$ varies smoothly (continuously differentiable) with position x and time t, and the initial density $k(x,0)$ is smooth with respect to position x, then the kinematic wave speed w will also vary smoothly with position x and time t, and the density $k(x,t)$ will vary smoothly along the path of the kinematic wave, such that $\frac{\mathrm{d}k(x,t)}{\mathrm{d}t} = -\frac{\partial \tilde{q}(k(x,t),x,t)}{\partial x}$.

However, if the road is homogeneous with respect to position x, i.e. $\frac{\partial \tilde{q}(k(x,t),x,t)}{\partial x} = 0$, then the flow, $q(x,t)$, density $k(x,t)$, and vehicle speed $v(x,t) = \frac{q(x,t)}{k(x,t)}$ from [2.4] will all remain constant along the path of the kinematic wave. In addition, if the road is also homogeneous with respect to time t, i.e. $\frac{\partial \tilde{q}(k(x,t),x,t)}{\partial t} = 0$, then the path of the kinematic wave is a straight line. The impacts of discontinuities in the equilibrium flow-density curve with respect to position x and time t, as well as the impacts of discontinuities in the initial density $k(x,0)$ are discussed in the following section.

2.3.6 Shockwaves at density discontinuities

A shockwave develops at any density discontinuity, which has a downstream density greater than the upstream density. Shockwaves may be propagated upstream with velocity $u > 0$ or downstream with velocity $u < 0$, or remain stationary with velocity $u = 0$. In reality a shockwave is spread over some finite region of the roadway where vehicles accelerate or decelerate from one flow state (defined by density and flow) to another. For simplification, a shockwave is often approximated using a single point which moves with the shockwave speed u. This simplification is equivalent to assuming that vehicles are capable of instantaneous acceleration and deceleration, which is consistent with the use of the equilibrium flow-density relationship.

The positive jump in density in the downstream direction ensures that the shock satisfies the entropy condition (Lax, 1973), which will be discussed in detail in the following section.

Richards (1956) showed that the shockwave speed is equal to the ratio of the difference in the flow to the difference in the density across the discontinuity, which is known as the Rankine-Huginot jump condition in fluid dynamics

[2.13] $$u(x_S,t) = \frac{q(x_2,t) - q(x_1,t)}{k(x_2,t) - k(x_1,t)} = \frac{\Delta q}{\Delta k}$$

where x_S is the position of the shockwave at time t and x_S is contained on the open interval (x_1, x_2).

[2.13] can be derived from a simple example. Consider a uniform stretch of road with flow state A downstream from flow state B, such that densities $k_A > k_B$, as depicted in Figure 2.5. Remember that the speed of the vehicles is equal to the slope of a radius vector from the origin to the relevant point on the flow-density curve, as shown in Figure 2.6. So, the faster-moving (less densely spaced) vehicles in state B catch up to the slower-moving (more densely spaced) vehicles in state A, creating a shockwave at the density discontinuity at position x_S. The position of the shockwave moves with speed $u(x_S,t)$ as shown in Figure 2.5, where the positive direction is from left to right. If $u(x_S,t) < 0$ then the shockwave would be moving from right to left.

Now, imagine two motorcycles driving with the shockwave speed $u(x_S,t)$ on the shoulder of the road spanning the shockwave such that the leader motorcycle is at position $x_r = \lim_{\varepsilon \to 0}(x_S + \varepsilon)$ and the follower motorcycle is at position $x_l = \lim_{\varepsilon \to 0}(x_S - \varepsilon)$. Vehicles pass the leader motorcycle at relative speed $v(x_r,t) - u(x_S,t)$. Applying the fundamental relationship [2.4], the flow of vehicles past the leader motorcycle is given by

$$\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{x_r} = k(x_r,t)\left[v(x_r,t) - u(x_S,t)\right],$$

which leads to

[2.14]
$$\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{x_r} = q(x_r,t) - k(x_r,t)u(x_S,t)$$

A similar equation applies to the follower motorcycle.

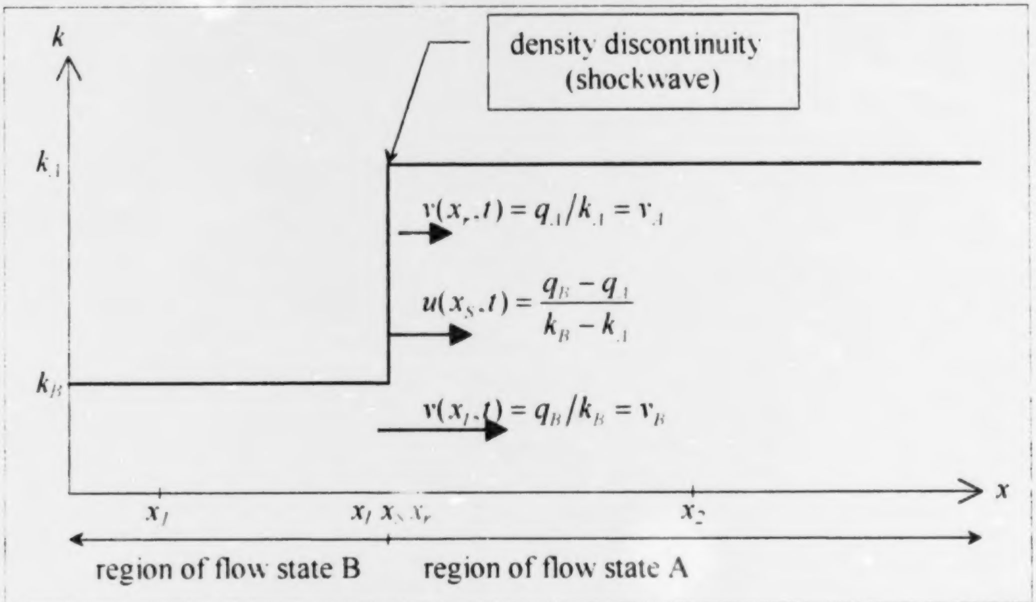

k
density discontinuity (shockwave)
k_A
$v(x_r,t) = q_A/k_A = v_A$
$u(x_S,t) = \frac{q_B - q_A}{k_B - k_A}$
k_B
$v(x_l,t) = q_B/k_B = v_B$
x
x_1
x_l x_S x_r
x_2
region of flow state B
region of flow state A

Figure 2.5 A shockwave at a density discontinuity

The rate at which vehicles enter the shockwave must equal the rate at which vehicles exit the shockwave. Therefore, the rate at which vehicles pass each motorcycle must be equivalent, i.e.

$$q(x_l,t) - k(x_l,t)u(x_S,t) = q(x_r,t) - k(x_r,t)u(x_S,t).$$

Changing the subscripts to either $l = 1$, $r = 2$ or $l = 2$, $r = 1$, and solving for $u(x_s,t)$ yields the shockwave speed [2.13].

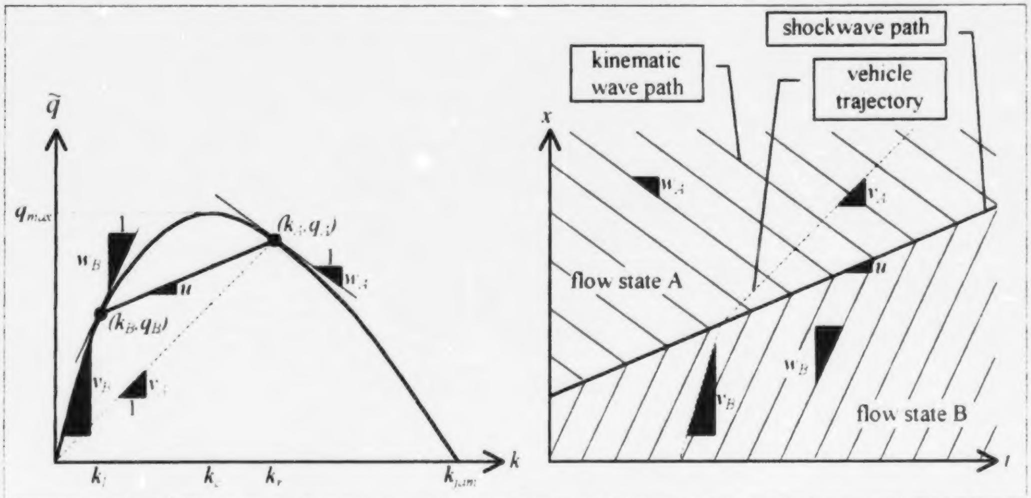

$\tilde{q}$
q_{max}
1
w_B
(k_A, q_A)
u
1
w_A
(k_B, q_B)
v_B
v_A
1
k_l
k_c
k_r
k_{jam}
k
shockwave path
kinematic wave path
vehicle trajectory
x
w_A
v_A
u
flow state A
w_B
v_B
flow state B
t

Figure 2.6 Relationship between the flow-density curve and the position-time graph for a discontinuous density along a homogeneous road

An alternative derivation (Lax, 1973) of [2.13] comes from fluid dynamics. Let $n(x_1, x_2, t)$ be the number of vehicles between positions x_1 and x_2 (see Figure 2.5) at time t

$$n(x_1, x_2, t) = \int_{x_1}^{x_2} k(x, t)\,\mathrm{d}x\,.$$

Taking the derivative with respect to t, then integrating by parts gives

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \int_{x_1}^{x_l} \frac{\partial k(x,t)}{\partial t}\,\mathrm{d}x + k(x_l, t)\frac{\mathrm{d}x_s}{\mathrm{d}t} + \int_{x_r}^{x_2} \frac{\partial k(x,t)}{\partial t}\,\mathrm{d}x - k(x_r, t)\frac{\mathrm{d}x_s}{\mathrm{d}t}\,.$$

Substituting $\frac{\mathrm{d}x_s}{\mathrm{d}t} = u$ by definition and $\frac{\partial k(x,t)}{\partial t} = -\frac{\partial q(x,t)}{\partial x}$ from conservation of vehicles [2.5] yields

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -\int_{x_1}^{x_l} \frac{\partial q(x,t)}{\partial x}\,\mathrm{d}x + k(x_l, t)u - \int_{x_r}^{x_2} \frac{\partial q(x,t)}{\partial x}\,\mathrm{d}x - k(x_r, t)u$$

$$\Rightarrow \frac{\mathrm{d}n}{\mathrm{d}t} = q(x_1, t) - q(x_l, t) + k(x_l, t)u + q(x_r, t) - q(x_2, t) - k(x_r, t)u$$

But it is clear from the conservation law that $\frac{dn}{dt} = q(x_1,t) - q(x_2,t)$. Equating the latter two equations and solving for the shockwave speed u simplifies to the Rankine-Huginot jump condition [2.13].

Graphically, the shockwave speed u is equal to the slope of the chord which joins the points on the flow-density curve that define the flow states upstream and downstream of the shockwave. This graphical rule has been applied to the example described in the above derivations. Figure 2.6 contains the flow-density diagram and the space-time diagram for the example. For known densities $k_l = k(x_l,t) = k_B$ and $k_r = k(x_r,t) = k_A$ the corresponding flow is determined using the equilibrium flow-density curve. The vehicle speed is the slope of the vector from the origin to the point on the equilibrium flow-density curve which defines the flow state. The kinematic wave speed is the slope of the tangent to the curve at the point defining the flow state. The speed of the shockwave is the slope of the vector which joins the points defining the flow states on either side of the shockwave. These slopes were used to draw the space-time diagram shown in Figure 2.6.

How do shockwaves affect flow, kinematic waves, and other shockwaves? It can be noted from [2.13] that any shockwave with a nonzero speed must be a discontinuity in flow, as well as a discontinuity in density. It is also evident from [2.13] that a stationary shockwave has a continuous flow across the density discontinuity. A kinematic wave ends when it meets a shockwave, and a new kinematic wave emerges on the other side of the shockwave. All shockwaves separate two flow states, hence the kinematic wave speed w, defined by [2.9], is discontinuous across the shockwave. Two kinematic waves which meet at a *stationary* shockwave happen to share the same flow. Nevertheless, they are still different kinematic waves with different wave speeds. When two shockwaves meet, if a new density discontinuity results between the adjacent flow states then a single new shockwave is generated.

The impact of the equilibrium flow-density relationship on shockwaves depends on the nature of the equilibrium flow-density relationship with respect to position x and time t. First, assume that a smooth (continuously differentiable) equilibrium flow-density curve $\tilde{q}(k(x,t))$ remains unchanged with respect to position x and time t. Any density discontinuity and its associated flow discontinuity are propagated with constant speed u given by [2.13]. Next, assume that the equilibrium flow-density curve $\tilde{q}(k(x,t),x,t)$ varies smoothly with position x and time t. Any density discontinuity and its associated flow discontinuity are propagated with speed $u(x,t)$ that may vary smoothly with position x and time t. Finally, if the equilibrium flow-density curve $\tilde{q}(k(x,t),x,t)$ depends discontinuously on position x, then this discontinuity acts as a stationary shockwave itself with continuous flow and discontinuous density across it. If the flow upstream of the stationary shockwave cannot be accommodated downstream, then a second backward, congestion-forming shockwave emanates from the stationary shockwave. Any other shockwaves which reach the stationary shockwave are absorbed and a new shockwave is generated with a speed that depends on the new density discontinuity as per [2.13].

It is interesting to note that non-smooth equilibrium flow-density curves, which includes most multi-regime models and all linear models, produce shockwaves even for a smooth initial density on a homogeneous roadway. At the point where the initial density crosses from one regime to another, convergent or divergent kinematic waves emerge, creating either a shockwave or a fan, respectively.

2.3.7 Analytical solution to the LWR hydrodynamic model

2.3.7.1 Smooth and weak solutions

Ansorge (1990) applied the work of Lax (1973) in fluid dynamics to demonstrate that the mathematical definition of the LWR hydrodynamic model is incomplete. Lax indicated that the conservation law specified in [2.5] does not allow for discontinuities in flow or density. Thus, the model is appropriate for problems that have continuously differentiable initial density such that the kinematic wave speed w increases with position x. Solutions to these initial value problems are called smooth solutions or genuine solutions. Smooth solutions satisfy [2.5] exactly, which implies that density and flow are continuous and differentiable for all positions and time $t \geq 0$.

To solve problems with densities that are not differentiable, Lax introduced weak solutions, also known as generalised solutions, which satisfy the conservation law in the sense of distribution theory, i.e. the conservation law is satisfied in its integral form. In this weak formulation of the LWR hydrodynamic model, weak solutions $k(x,t)\ \forall x \in \Re, t \in \Re^+$ satisfy

$$-\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left[k\frac{\partial \phi}{\partial t}+q\frac{\partial \phi}{\partial x}\right]\mathrm{d}t\,\mathrm{d}x-\int_{-\infty}^{+\infty}\phi(x,0)k(x,0)\,\mathrm{d}x=0 \qquad \forall \phi \in \{x \in \Re, t \in \Re_0^+\}.$$

which is obtained by multiplying [2.5] by test function $\phi(x,t)$, integrating the resulting equation, then integrating by parts. The test function $\phi(x,t)$ is defined as any continuously differentiable function on the half space $\{x \in \Re, t \in \Re^+\}$ with compact support on $\{x \in \Re, t \in \Re_0^+\}$.

All smooth solutions are contained within the set of weak solutions. The weak solutions, which have continuously differentiable density, are smooth solutions. Weak solutions, which are not smooth solutions, permit density discontinuities provided that each such discontinuity satisfies the Rankine-Huginot jump condition that defines the shockwave speed [2.13]. Note that the weak formulation of the LWR hydrodynamic model and the Rankine-Huginot jump condition were applied graphically to arrive at the solution in Figure 2.6. These graphical rules will be explained later.

The LWR hydrodynamic model may be used to solve initial value problems, in which the future density on the road must be found, given an initial density on the road. Consider an initial value problem in which the initial density has a discontinuity such that the downstream density is smaller. Assume that the downstream flow state is (k_B, q_B) and the upstream state is (k_A, q_A), where flow states A and B are the same as previously defined. For this initial value problem two weak solutions, which differ only in the grey cone, are depicted in Figure 2.7.

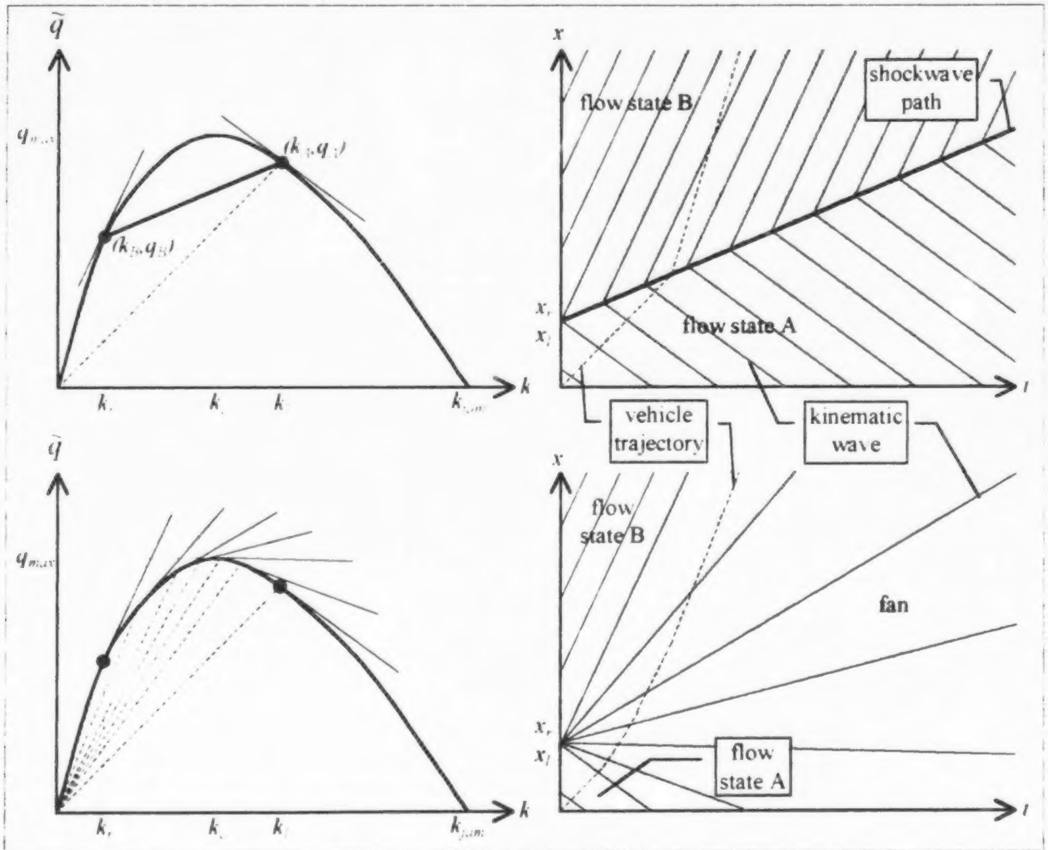

$\tilde{q}$
q_{max}
(k_A, q_A)
(k_B, q_B)
k_r
k_c
k_l
k_{jam}
k
x
flow state B
shockwave path
x_r
x_l
flow state A
t
vehicle trajectory
kinematic wave
flow state B
fan
flow state A

Figure 2.7 Two weak solutions of the LWR hydrodynamic model for a density discontinuity

Note in Figure 2.7 that the positions of k_l and k_r have been switched compared to Figure 2.6.

The upper weak solution in Figure 2.7 contains a shockwave which is an abrupt change in density. The lower weak solution contains a gradual density transition by following the equilibrium flow-density curve. The paths of the kinematic waves on the space-time diagram fan out from the initial density discontinuity, as though the density discontinuity were replaced by a rapid, smooth decline in density. Observation of real traffic under similar conditions will verify that the physically relevant solution is the lower weak solution. The presence of multiple weak solutions begs the question: How can the physically relevant solution be identified without having to observe traffic? The answer to this question is the introduction of an entropy maximising condition.

2.3.7.2 The entropy condition and driver's ride impulse

The weak formulation of the LWR hydrodynamic model permits infinitely many weak solutions, of which only one is physically relevant. Lax (1971) recommended a completion of the LWR hydrodynamic model by adding the entropy condition. The second main theorem of thermodynamics states that $\mathrm{dE} \geq 0$, where E is the physical entropy. The entropy condition was restated (Lax, 1973) in the form of kinematic waves and shockwaves as

$$w(x_l,t) > u(x_s,t) > w(x_r,t) \tag{2.15}$$

where flow is from left (l) to right (r). The expression of the entropy condition assumes a continuously differentiable, concave equilibrium flow-density relationship. On a space-time diagram the entropy condition implies that kinematic waves on both sides of the discontinuity must intersect the shockwave with increasing time t. A shockwave can now be more specifically defined (Lax, 1973) as a density discontinuity that satisfies the Rankine-Huginot jump condition [2.13] and the entropy condition [2.15]. Referring to Figure 2.7, the shockwave in the upper weak solution does not satisfy the entropy condition, so the upper weak solution is not an entropy solution of the completed LWR hydrodynamic model.

Ansorge (1990) reinterpreted the entropy condition as driver's ride impulse

[2.16] $$k(x_l,t) \le k(x_r,t)$$

[2.16] implies that drivers either (1) tend to increase their density instantaneously as they cross a discontinuity in the flow-state, or (2) smooth the discontinuity so that the density profile has no discontinuity. Case (1) corresponds to a shockwave in which the upstream density must be less than the downstream density, i.e. $k(x_l,t) < k(x_r,t)$. Case (2) produces a fan of characteristics on which the density profile has no discontinuities, hence $k(x_l,t) = k(x_r,t)$. The difference between a shockwave and a fan is described in more detail in the following section.

If the equilibrium flow-density curve is concave and differentiable, then a shockwave satisfies the driver's ride impulse $k(x_l,t) < k(x_r,t)$ if and only if the entropy condition is satisfied. The proof is as follows.

The concavity of the equilibrium flow-density relationship can be stated mathematically as

[2.17] $$\tilde{q}(k_1) < \tilde{q}(k_2) + \tilde{q}_k(k_2)(k_1 - k_2) \quad \forall\, k_1, k_2 \in [0, k_{jam}],\ k_1 \ne k_2$$

where subscript $_k$ indicates a partial derivative with respect to density.

Substitution of $k_1 = k(x_r,t) = k_r$ and $k_2 = k(x_l,t) = k_l$ into [2.17] leads to

[2.18] $$\tilde{q}(k_r) - \tilde{q}(k_l) < \tilde{q}_k(k_l)(k_r - k_l) \quad \forall\, k_r, k_l \in [0, k_{jam}],\ k_r \ne k_l$$

The sufficient condition can be proved by contradiction. The entropy condition states that $u < w(k_l,t)$. Substitution of the Rankine-Huginot jump condition [2.13] on the left hand side, and the kinematic wave speed [2.10] on the right hand side, yields

$$\frac{\tilde{q}(k_r)-\tilde{q}(k_l)}{k_r-k_l}<\tilde{q}_k(k_l).$$

The hypothesis $k_r - k_l < 0$ implies that

$$\tilde{q}(k_r)-\tilde{q}(k_l)>\tilde{q}_k(k_l)(k_r-k_l),$$

which contradicts the assumption of concavity [2.18]. Therefore, the hypothesis is false, and it can be concluded that the driver's ride impulse is true $k_l \le k_r$. Obviously, if $k_l = k_r$ then there was no initial density discontinuity, so for shockwaves $k_l < k_r$. Thus the sufficient condition has been proven.

Now, consider the necessary condition to complete the proof. [2.17] with $k_1 = k(x_l, t) = k_l$ and $k_2 = k(x_r, t) = k_r$, leads to

$$\tilde{q}(k_r)-\tilde{q}(k_l)>\tilde{q}_k(k_r)(k_r-k_l) \quad \forall\, k_l, k_r \in [0, k_{jam}],\ k_l \ne k_r \tag{2.19}$$

Since $k_l \ne k_r$, driver's ride impulse implies that $k_l < k_r$. Therefore, [2.19] leads to

$$\frac{\tilde{q}(k_r)-\tilde{q}(k_l)}{k_r-k_l}>\tilde{q}_k(k_r)\Rightarrow u>\tilde{q}_k(k_r).$$

Similarly, [2.18] with $k_l < k_r$ leads to

$$\frac{\tilde{q}(k_r)-\tilde{q}(k_l)}{k_r-k_l}<\tilde{q}_k(k_l)\Rightarrow u<\tilde{q}_k(k_l).$$

So, driver's ride impulse implies that $\tilde{q}_k(k_l) > u > \tilde{q}_k(k_r)$, which is consistent with the entropy condition. This completes the proof.

The entropy condition and driver's ride impulse can be interpreted graphically. An arrow is drawn on the flow-density plot (Figure 2.8), which points from the upstream flow state $(k_l, \tilde{q}(k_l))$ to the downstream flow state $(k_r, \tilde{q}(k_r))$. All shockwaves must have an arrow pointing to the right, representing a discontinuous jump in density which is consistent with [2.15] and [2.16]. All arrows which point to the left indicate the presence of a fan, in which all states on the equilibrium flow-density curve between the upstream density k_l and the downstream density k_r occur. This is consistent with driver's ride impulse [2.16].

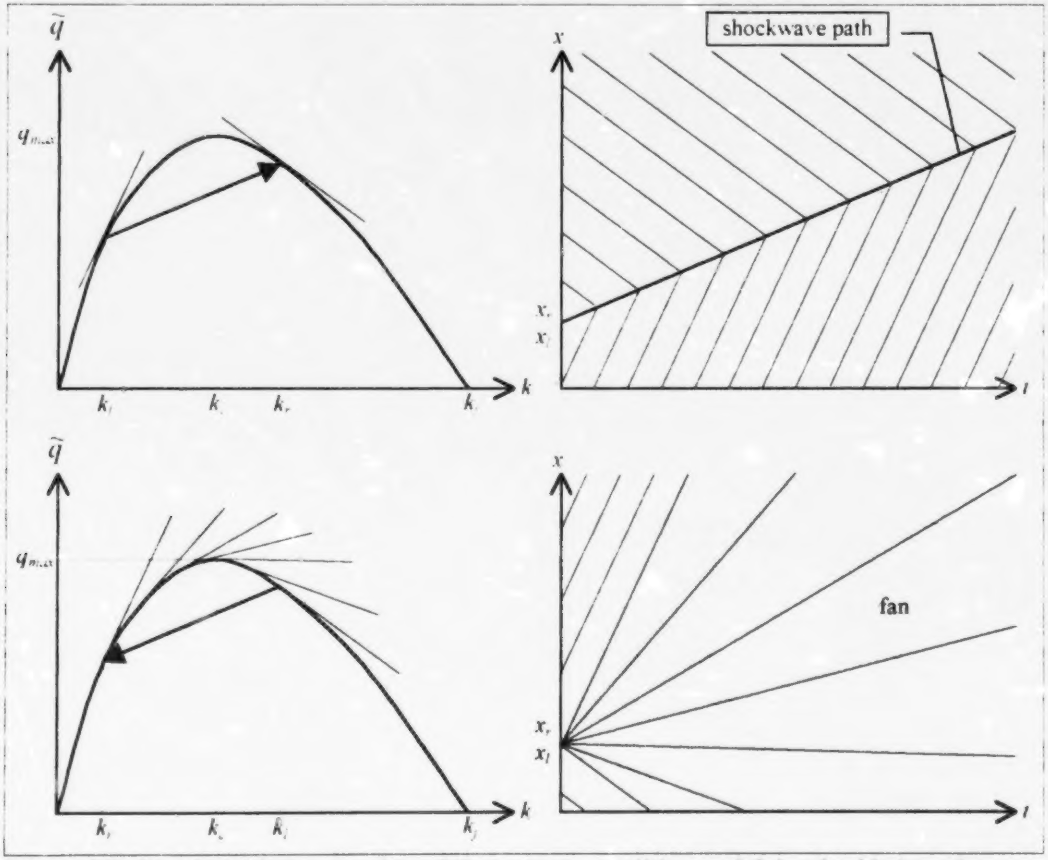

$\tilde{q}$
q_{max}
k_l
k_c
k_r
k_j
k
x
shockwave path
x_r
x_l
t
$\tilde{q}$
q_{max}
k_r
k_c
k_l
k_j
k
x
fan
x_r
x_l
t

Figure 2.8 Graphical interpretation of the entropy condition and driver's ride impulse

2.3.7.3 Unique solution to the LWR hydrodynamic model

For the case of a non-linear, differentiable equilibrium flow-density relationship which is independent of position x and time t, Lax (1973) proved that every initial value problem has exactly one weak solution which has only shockwaves as discontinuities, and the shockwaves satisfy the Rankine-Huginot jump condition [2.13] as well as the entropy condition [2.15]. The entropy condition applies only to cases in which the equilibrium flow-density relationship is concave and continuously differentiable. A more general proof has not been developed for the case of a piecewise linear equilibrium flow-density relationship, or for the case of an equilibrium flow-density relationship that may vary with position x and/or time t.

Some light is shed on the implications of a piecewise linear equilibrium flow-density relationship on the unique solution of the LWR model by applying the graphical interpretation of the entropy condition. Consider a triangular equilibrium flow-density relationship. Any density in the uncongested regime, i.e. $k_l, k_r < k_c$, would cause a shockwave to propagate at the free speed, regardless of the magnitude or sign of the density discontinuity. Likewise, any density in the congested regime, i.e. $k_l, k_r > k_c$, would cause a shockwave to propagate at the backward wave speed defined by $w = \tilde{q}_k(k), k > k_c$, regardless of the magnitude or sign of the density discontinuity.

Any density discontinuity, which does not fall into one of the two categories above, must span the critical density with either a positive or negative jump in density. For the positive jump in density, i.e. $k_l < k_c < k_r$, the graphical rule of Figure 2.8 dictates that the density discontinuity is a regular shock, as shown in the upper diagram of Figure 2.9. For the negative jump in density, i.e. $k_l > k_c > k_r$, the graphical rule of Figure 2.8 suggests that a fan would develop that represents all the flow states on the interval $[k_r, k_l]$. The fan does not spread smoothly across the cone because all of the characteristics for flow states with density $k \in (k_r, k_c)$ have the same slope as the characteristic with density $k = k_r$, and all of the characteristics for flow states with density $k \in (k_c, k_j)$ have the same slope as the characteristic with density $k = k_j$. So the characteristics hug the edges of the cone, as shown in the lower part of Figure 2.9. These overlapping characteristics on the edges of the cone are equivalent to shockwaves because they represent a density discontinuity that propagates with speed given by the Rankine-Huginot jump condition [2.13]. Between the shockwaves is a cone that has flow state (k_c, q_{max}).

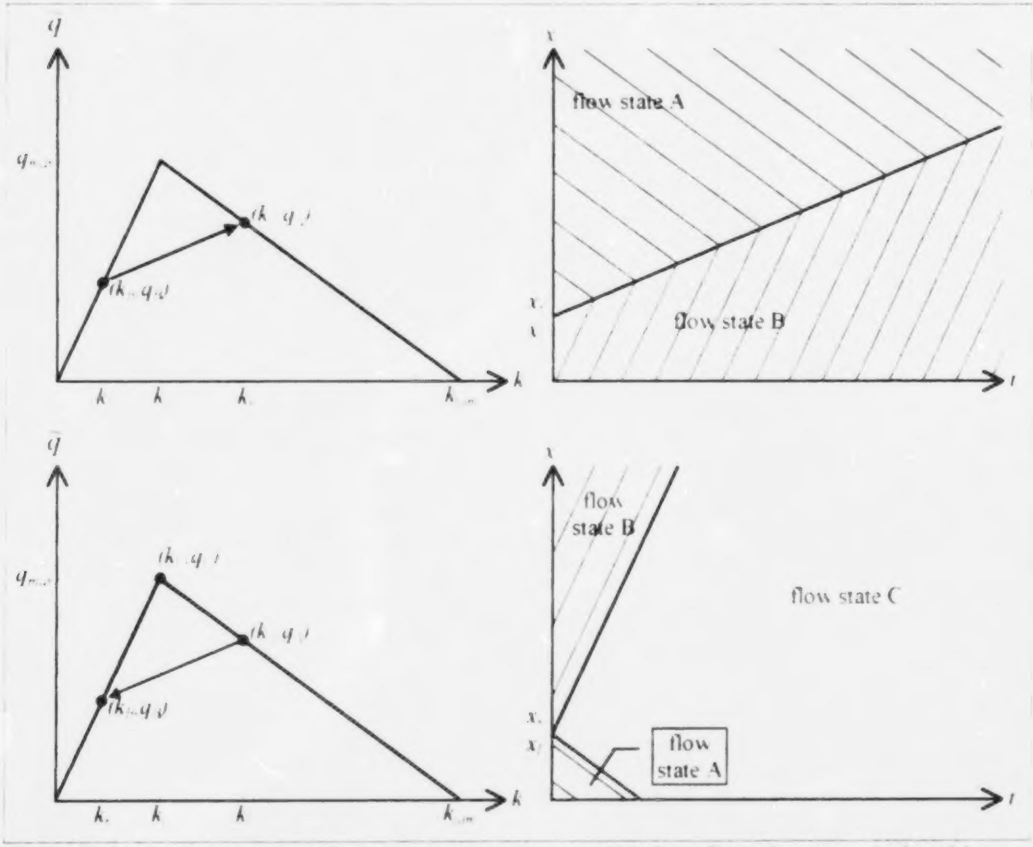

q
q_max
(k, q)
(k, q)
k
k
k
k
k_jam
k
x
flow state A
flow state B
x
x
t
q
q_max
(k, q)
(k, q)
(k, q)
k
k
k
k_jam
k
x
flow
state B
flow state C
x
x
flow
state A
t

Figure 2.9 Shockwaves and fans for a triangular equilibrium flow-density relationship

The changing density profile with time is shown in Figure 2.10 for a negative jump in density. Note that the width of the profile line decreases with time. The upper diagram is the resulting step density profile for the diverging shockwaves discussed above, which result from a triangular equilibrium flow-density relationship. The lower diagram shows the spreading of a fan that occurs if the equilibrium flow-density relationship is continuously differentiable. The shockwaves spread faster than the fan because they travel at the free-speed and the backward wave speed.

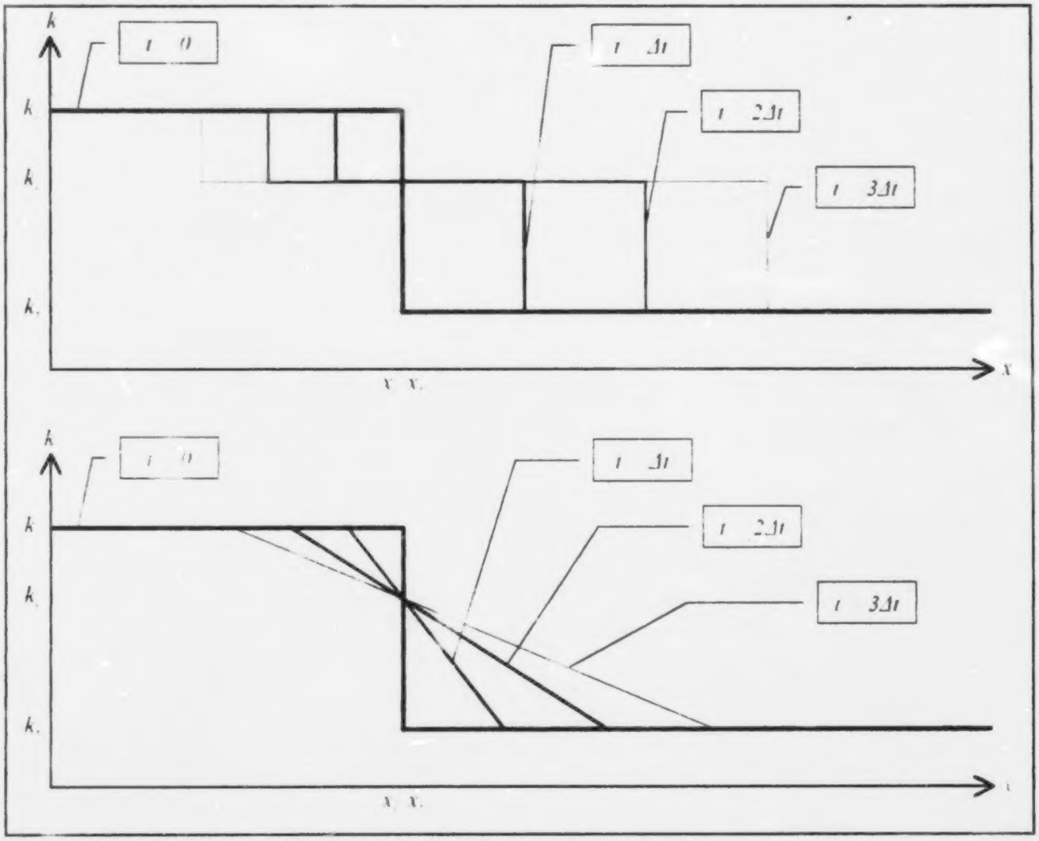


Figure 2.10 Density profile for an initial density discontinuity

The piecewise-linear equilibrium flow-density relationship can be considered as an approximation of the non-linear equilibrium flow-density relationship. The approximation is improved as more linear pieces, and consequently more non-differentiable points, are added. At each non-differentiable point that lies on the interval $[k_r, k_l]$, a shock forms which adds another step to the density profile. With an infinite number of linear functions and non-differentiable points, the equilibrium flow-density relationship becomes continuously differentiable, and the step profile becomes a smooth fan.

Therefore, the LWR model with a piecewise-linear flow-density relationship has a unique weak solution. So, the entropy condition is no longer necessary to distinguish the physically relevant solution from other weak solutions. This finding is convenient because the shockwaves that form the steps in the density profile in Figure 2.10 do not respect the entropy condition. Indeed, all the shocks discussed above with $k_l > k_r$ violate driver's ride impulse as well. Therefore, neither the entropy condition nor driver's ride impulse applies if the flow-density relationship is piecewise-linear.

2.3.7.4 Solution by the method of characteristics

Initially proposed by Lighthill and Whitham (1955) and Richards (1956), the method of characteristics was succinctly described by Daganzo (1995b).

A characteristic is a curve in the space-time plane emanating from a known point in space-time defined by the initial conditions. The state of traffic at any point in space-time can be traced back to a single point of the initial conditions, and is independent of the initial conditions at any other point.

A characteristic at any point follows the path of the kinematic wave at that point. Therefore, the flow along the characteristic is constant. As noted in the discussion of kinematic waves, if the equilibrium flow-density relationship is independent of position then the density is also constant along the characteristic, and if the equilibrium flow-density relationship is independent of time then the characteristic traces a straight line in time-space.

Density discontinuities in the initial conditions can be propagated in the space-time plane using shockwaves and/or divergent characteristics, which are also called fans. Assuming a concave equilibrium flow-density curve, as is usually the case, an increase in the density across the discontinuity in the downstream direction produces a shockwave, as discussed above. A kinematic wave ends at a shockwave, and another kinematic wave with a different speed emerges from the other side of the shockwave. A characteristic which meets a shockwave will cross the shockwave, and follow the path of the kinematic wave on the other side of the shockwave with whatever flow is present on the new kinematic wave. A decrease in the density across the discontinuity creates a divergent fan of characteristics, just as a fan of kinematic waves would diverge from a rapid, smooth change in the initial density. If the equilibrium flow-density curve is convex in certain regions, then a combination of fans and shockwaves may be produced at the initial density discontinuity.

A discontinuity in the equilibrium flow-density relationship behaves like a stationary shockwave, as described in the previous discussion of shockwaves. A characteristic which reaches this stationary shockwave will have a discontinuous change in speed.

2.3.8 Finite difference approximations to the LWR hydrodynamic model

Luke (1972) and Newell (1993) proposed exact solutions for certain classes of kinematic wave problems with particular flow-density relationships. For more general problems, however, the analytical solution is too cumbersome, and mechanisation of the procedure for a computer programme is very difficult. Therefore, approximations to the LWR model were developed that were suitable for implementation into a computer program. After traditional numerical methods are introduced, two other approximations, both of which are related, are documented in the following sections.

2.3.8.1 Numerical methods

A brief discussion of the application of numerical methods to solve hyperbolic systems of conservation laws, such as the LWR hydrodynamic model, is given by Lax (1973). The space-time plane is discretised into a mesh defined by lattice points (x_i, t_j) where i and j are integers. The density is specified at these lattice points as $k_i^j = k(x_i, t_j)$. The density at time t_j is shown in Figure 2.11 with density values at lattice points indicated by dots.

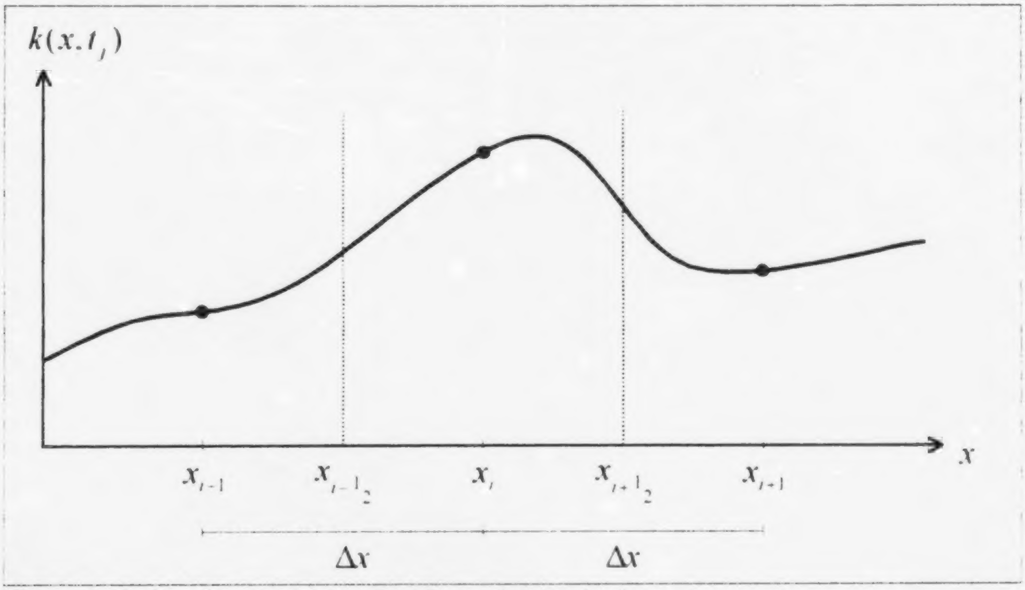


Figure 2.11 Density at lattice points for an instant in time

If it is assumed that the density k_i^j is an approximation to the average density over the cell $\left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right]$ at time t_j, then the conservation law implies

[2.20]
$$k_i^{j+1} = k_i^j + \frac{\Delta t}{\Delta x}\left(\overline{q}_{i-\frac{1}{2}}^j - \overline{q}_{i+\frac{1}{2}}^j\right).$$

where $\overline{q}_i^j = \frac{1}{\Delta t}\int_{t_j}^{t_{j+1}} \widetilde{q}(k(x_i, t))\,\mathrm{d}t$ is the average equilibrium flow at position x_i during the time interval $[t_j, t_{j+1}]$. Approximating $\overline{q}_{i+\frac{1}{2}}^j$ with

[2.21] $$\bar{q}^{t}_{i+1/2} = \frac{\tilde{q}(k^{t}_{i+1}) + \tilde{q}(k^{t}_{i})}{2} + \frac{\Delta x}{\Delta t}\frac{k^{t}_{i} - k^{t}_{i+1}}{2}$$

leads to the crude differencing scheme proposed by Lax (1954):

$$k_{i}^{t+1} = \frac{k^{t}_{i+1} + k^{t}_{i-1}}{2} + \frac{\Delta t}{\Delta x}\frac{\tilde{q}(k^{t}_{i-1}) - \tilde{q}(k^{t}_{i+1})}{2}$$

In words, the occupancy of a cell at the next time step is the average of the occupancy of the upstream and downstream cells at the present time, plus the number of vehicles which enter the cell during one time step at the present flow rate, minus the number of vehicles which exit the cell during one time step at the present flow rate. There are several sources of error in this discretisation scheme. Firstly, the flow rate at the cell boundary is approximated using [2.21]. The actual flow rate between the cells depends on the density at the cell boundary, which is assumed to be an average of the density at the cell midpoints. The actual flow rate changes during the time step. Secondly, the future density at the midpoint of a cell depends equally on the upstream and downstream cells, and not at all on the present density of the cell itself. Finally, the cell density is known only at the midpoint of the cell. Cell occupancy is estimated from interpolations based on these values. Nevertheless, this crude differencing scheme was proven to converge to the exact entropy solution of the LWR model for finer discretisations of space-time (Vvedenskaya, 1956).

Any discretisation scheme of space-time must respect the CFL condition (Courant, Friedrichs and Lewy, 1928) to ensure stability

[2.22] $$\frac{\Delta x}{\Delta t} \ge \max_{k} \frac{\partial \tilde{q}}{\partial k}.$$

Numerical error is introduced into the solution in the form of artificial viscosity, which tends to smear the density with each time step, rendering shockwaves obscure. The magnitude of the error is proportional to $\frac{(\Delta x)^2}{\Delta t}\frac{\partial^2 k}{\partial x^2}$. So, the numerical error due to artificial viscosity can be minimised by choosing the space discretisation such that

$$\Delta x = \Delta t \cdot \max_k \frac{\partial \tilde{q}}{\partial k}. \tag{2.23}$$

This numerical approximation can be improved with a more accurate approximation to $\bar{q}^{j}_{i+\frac{1}{2}}$.

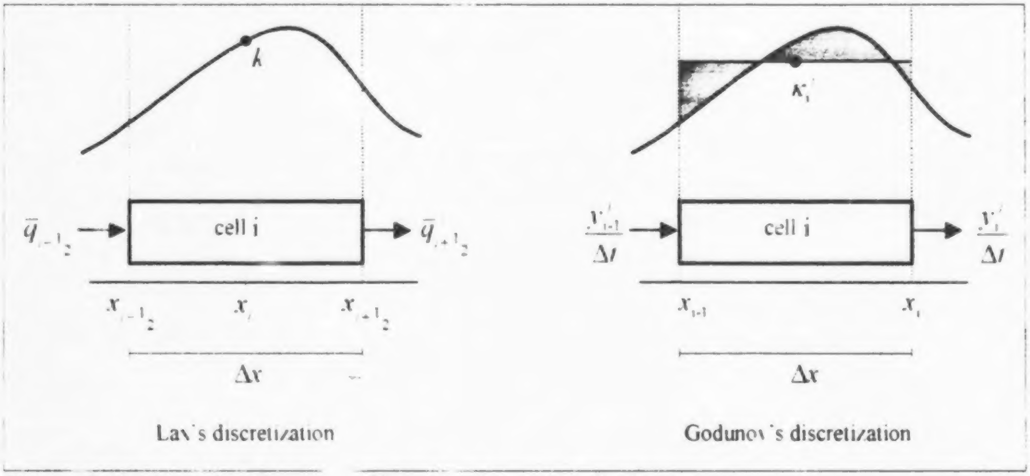
For example, Richtmyer and Morton (1967) suggested replacing [2.21] with an approximation to the average equilibrium flow which is centred on the time interval $[t_j, t_{j+1}]$

$$\begin{aligned} &\bar{q}^{j}_{i+\frac{1}{2}} = \tilde{q}\left(\bar{k}^{j+\frac{1}{2}}_{i+\frac{1}{2}}\right) \\ &\text{where } \bar{k}^{j+\frac{1}{2}}_{i+\frac{1}{2}} = \frac{k_i^j + k_{i+1}^j}{2} + \frac{\Delta t}{\Delta x}\frac{\tilde{q}(k_i^j) - \tilde{q}(k_{i+1}^j)}{2}. \end{aligned} \tag{2.24}$$

2.3.8.2 Godunov approximate solution

Another type of difference scheme, which resolves non-linear hyperbolic conservation equations, was introduced by Godunov (1959). [2.20] is employed, but the density k_i^j is first replaced with the average density κ_i^j of cell i. Using a piecewise-constant approximation to the density allows for the exact determination of the average equilibrium flows as the solution to Riemann problems.

The notation for the space discretisation and cell placement differs from the previous section as displayed in Figure 2.12. Cell i spans the region $[x_{\mathrm{i}-1}, x_{\mathrm{i}}]$, which is equivalent to the previously defined $\left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right]$ (note italics in subscript). Instead of approximating the cell density with the central density k_i^j, the average density κ_i^j is used. The number of vehicles which exit cell i during the time interval $[t_j, t_{j+1}]$ is given by y_i^j. The general solution algorithm to the initial value problem with a hyperbolic conservation equation is described in Figure 2.13.



k
$\bar{q}_{i-1/2}$
cell i
$\bar{q}_{i+1/2}$
$x_{i-1/2}$
x_i
$x_{i+1/2}$
Δx
Lax's discretization
κ_i^j
$\frac{y_{i-1}^j}{\Delta t}$
cell i
$\frac{y_i^j}{\Delta t}$
x_{i-1}
x_i
Δx
Godunov's discretization

Figure 2.12 Notation for space discretisation and cell placement

Figure 2.13 General solution algorithm of the Godunov scheme

Step 0. Initialisation

The road is discretised into cells. Each cell i covers the road section $[x_{i-1}, x_i]$ with length $\Delta x_i = x_i - x_{i-1}$. Time is discretised into time intervals $[t_j, t_{j+1}]$ each of length Δt. Initialise the iteration $j = 0$.

Step 1. Approximate the actual density with a piecewise constant density

Replace the actual density $k(x, t_j)$ by a piecewise constant density

$$\kappa_i^j = \kappa(x, t_j) = \frac{1}{\Delta x_i} \int_{x_{i-1}}^{x_i} k(x, t_j)\mathrm{d}x \quad \forall x \in [x_{i-1}, x_i] \ \forall i$$

so that every position in each cell has the average density for that cell.

Step 2. Calculate the exact density

Find the exact density $k(x, t_{j+1})$ $\forall i$ by solving

$$\frac{\partial k}{\partial t} + \frac{\partial (q(k))}{\partial x} = 0$$
$$k(x,0) = \kappa_i^j \quad \forall x \in [x_{i-1}, x_i]$$

Step 3. Stopping criterion

If $t_{j+1} \le \bar{t}$ where $\bar{t}$ is the end of the simulation, then increment the iteration $j \leftarrow j+1$ and return to Step 1. Otherwise, stop.

Alternatively, Lebacque (1996) suggested finding the piecewise-constant density κ_i^{j+1} $\forall$i directly using a two step process, which avoids the repetition of the piecewise-constant approximation of density. Hence, only the initial density is transformed into a piecewise-constant density during the initialisation. The modified algorithm is as shown in Figure 2.14.

Figure 2.14 Modified solution algorithm of the Godunov scheme

Step 0. Initialisation

The road is discretised into cells. Each cell i covers the road section $[x_{i-1}, x_i]$ with length $\Delta x_i = x_i - x_{i-1}$. Time is discretised into time intervals $[t_j, t_{j+1}]$ each of length Δt. Initialise the iteration $j = 0$. Replace the actual density $k(x, t_j)$ by a piecewise constant density

$$\kappa_i^j = \kappa(x, t_j) = \frac{1}{\Delta x_i} \int_{x_{i-1}}^{x_i} k(x, t_j)\mathrm{d}x \quad \forall x \in [x_{i-1}, x_i] \ \forall i$$

so that every position in each cell has the average density for that cell.

Step 1. Calculate flow

Find the number of exiting vehicles y_i^j which exit cell i by crossing position x_i during the time interval $[t_j, t_{j+1}]$ using

$$y_i^j = \int_{t_j}^{t_{j+1}} q(k(x_i, t))\,\mathrm{d}t \quad \forall \mathrm{i}.$$

Step 2. Calculate density

Update the piecewise-constant density with

$$\kappa_i^{j+1} = \kappa_i^j + \frac{y_{i-1}^j - y_i^j}{\Delta x_i} \quad \forall i.$$

Step 3. Stopping criterion

If $t_{j+1} \le \bar{t}$ where $\bar{t}$ is the end of the simulation, then increment the iteration $j \leftarrow j+1$ and return to Step 1. Otherwise, stop.

Lebacque (1996) remarked that Step 1 is a generalised Riemann problem at each cell boundary. A Riemann problem is a density discontinuity which separates two flow states each of constant density with respect to position. The solution to the generalised Riemann problem is the flow at the density discontinuity. Usually, the Riemann problem assumes that the equilibrium flow-density curve is constant with respect to position and time. However, in a generalised Riemann problem the equilibrium flow-density curve is permitted to vary spatially and/or temporally, continuously or discontinuously. Figure 2.15 presents the case in which the equilibrium flow-density relationship is discontinuous at the cell boundary, but homogeneous within cells.

A simple and efficient solution algorithm for the generalised Riemann problem would make the Godunov scheme very attractive for implementation within a computer program. Lebacque (1996) applied the graphical methods previously described to exhaustively enumerate the solutions for 18 distinct cases which collectively describe all possible configurations of the generalised Riemann problem. These 18 cases can be conveniently summarised by the four possible combinations of congested and uncongested conditions upstream (left) and downstream (right) of the cell boundary. The results are shown in Table 2.1.

Case 1. Upstream and downstream uncongested: The characteristics travel downstream on both sides of the cell boundary, so the flow on the upstream side of the boundary controls the solution the Riemann problem. The upstream flow crosses the cell boundary, provided that the downstream cell can accept that flow. Recall that the discontinuity in the equilibrium flow-density relationship acts like a stationary shockwave, at which the change in flow across the discontinuity must be zero, according to the Rankine-Huginot jump condition [2.13]. Since the downstream cell is uncongested, the flow entering cell i + 1 may be as large as the maximum flow $\tilde{q}_{\max,i+1} = \tilde{q}_{i+1}(k_{c,i+1})$. Hence, the solution to the generalised Riemann problem for this case is given by $q(x_i, t_j) = \min(\tilde{q}_i(\kappa_i), \tilde{q}_{\max,i+1})$.

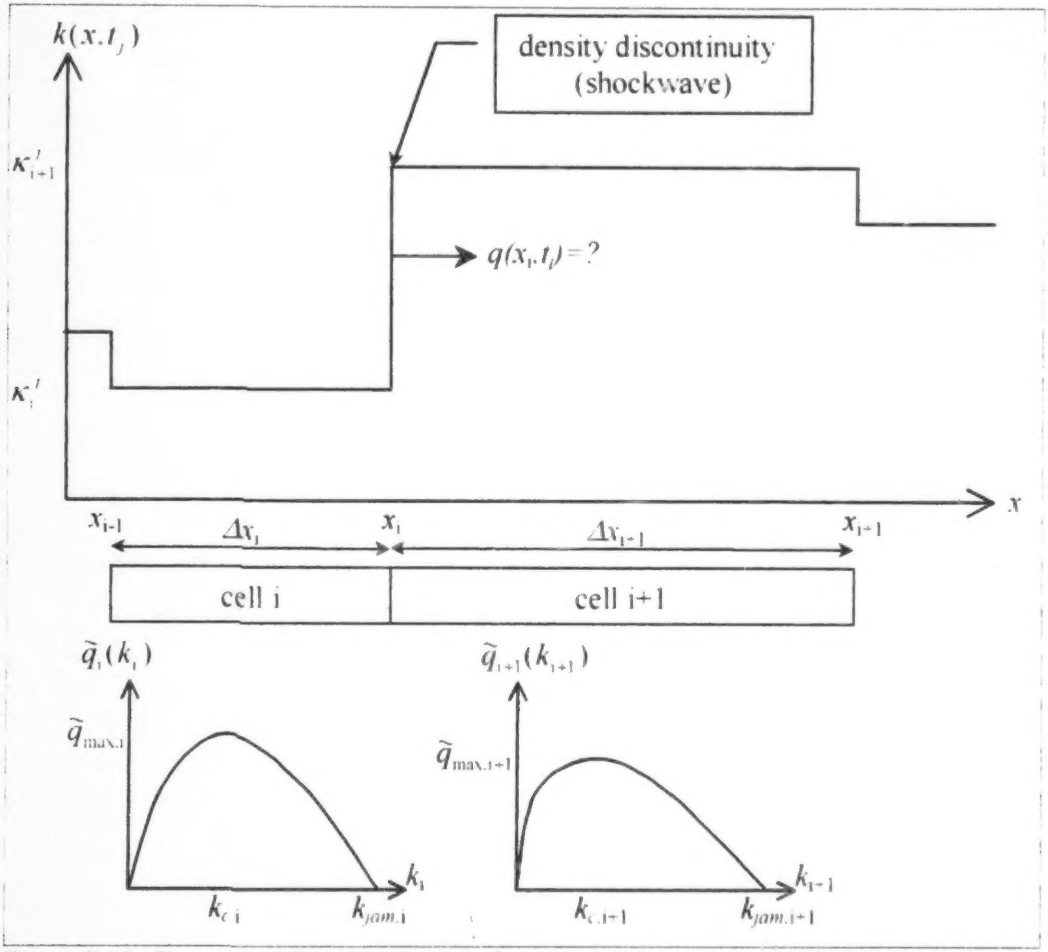

$k(x,t_j)$
density discontinuity (shockwave)
κ^j_{i+1}
$q(x_i,t_j) = ?$
κ^j_i
x
x_{i-1}
Δx_i
x_i
Δx_{i+1}
x_{i+1}
cell i
cell i+1
$\tilde{q}_i(k_i)$
$\tilde{q}_{\max,i}$
k_i
$k_{c,i}$
$k_{jam,i}$
$\tilde{q}_{i+1}(k_{i+1})$
$\tilde{q}_{\max,i+1}$
k_{i+1}
$k_{c,i+1}$
$k_{jam,i+1}$

Figure 2.15 Generalised Riemann problem with discontinuous equilibrium flow-density at cell boundary

Case 2. Upstream uncongested and downstream congested: The characteristics travel towards the cell boundary from both sides of the cell boundary. As in Case 1, the upstream flow will cross the cell boundary, provided that the downstream cell can accept that flow. Since the downstream cell is congested, the flow entering cell $k_i = k_i$ may be only as large as the current downstream flow. Therefore, the solution to the Riemann problem is the smaller of the two flows, i.e.

$q(x_i, t_j) = \min\left(\tilde{q}_i(\kappa_i), \tilde{q}_{i+1}(\kappa_{i+1})\right)$.

Case 3. Upstream congested and downstream uncongested: The characteristics travel away from the cell boundary on both sides of the cell boundary. The upstream congestion will recover as the flow pours out of the upstream cell at the maximum upstream flow $\tilde{q}_{\max,i} = \tilde{q}_i(k_{c,i})$, provided that the downstream cell can accommodate such a flow. Because the downstream cell is uncongested, the recovery will be limited by the maximum downstream flow, as in Case 1. So,

$q(x_i, t_j) = \min\left(\tilde{q}_{\max,i}, \tilde{q}_{\max,i+1}\right)$

Case 4. Upstream and downstream congested: The characteristics travel upstream on both sides of the cell boundary, so the downstream congestion will control the solution to the Riemann problem. The upstream congestion pours across the cell boundary at the maximum upstream flow, provided that the downstream cell can accommodate such a flow. If not, the recovery will be limited by the downstream flow, as in Case 3. So, $q(x_i, t_j) = \min\left(\tilde{q}_{\max,i}, \tilde{q}_{i+1}(\kappa_{i+1})\right)$.

Table 2.1 Solution to generalised Riemann problem $q(x_i, t_j)$ by exhaustive enumeration

	cell i + 1 uncongested $\kappa_{i+1}^{t} \le k_{c,i+1}$	cell i + 1 congested $\kappa_{i+1}^{t} \ge k_{c,i+1}$
cell i uncongested $\kappa_i^{t} \le k_{c,i}$	Case 1 $\min\left(\tilde{q}_i(\kappa_i), \tilde{q}_{\max,i+1}\right)$	Case 2 $\min\left(\tilde{q}_i(\kappa_i), \tilde{q}_{i+1}(\kappa_{i+1})\right)$
cell i congested $\kappa_i^{t} \ge k_{c,i}$	Case 3 $\min\left(\tilde{q}_{\max,i}, \tilde{q}_{\max,i+1}\right)$	Case 4 $\min\left(\tilde{q}_{\max,i}, \tilde{q}_{i+1}(\kappa_{i+1})\right)$

Daganzo (1995b) further simplified the solution to the generalised Riemann problem by splitting the equilibrium flow-density relationship into two curves: local demand and local supply. The local demand curve is the uncongested portion of the equilibrium flow-density curve, extended through the congested regime at the maximum flow

[2.25] $$q_D(k) = \begin{cases} \tilde{q}(k) & \text{if } k \le k_c \text{ (uncongested regime)} \\ \tilde{q}(k_c) & \text{if } k \ge k_c \text{ (congested regime)} \end{cases}$$

The local supply curve is the congested portion of the equilibrium flow-density curve, extended through the uncongested regime at the maximum flow

[2.26] $$q_S(k) = \begin{cases} \tilde{q}(k_c) & \text{if } k \le k_c \text{ (uncongested regime)} \\ \tilde{q}(k) & \text{if } k \ge k_c \text{ (congested regime)} \end{cases}$$

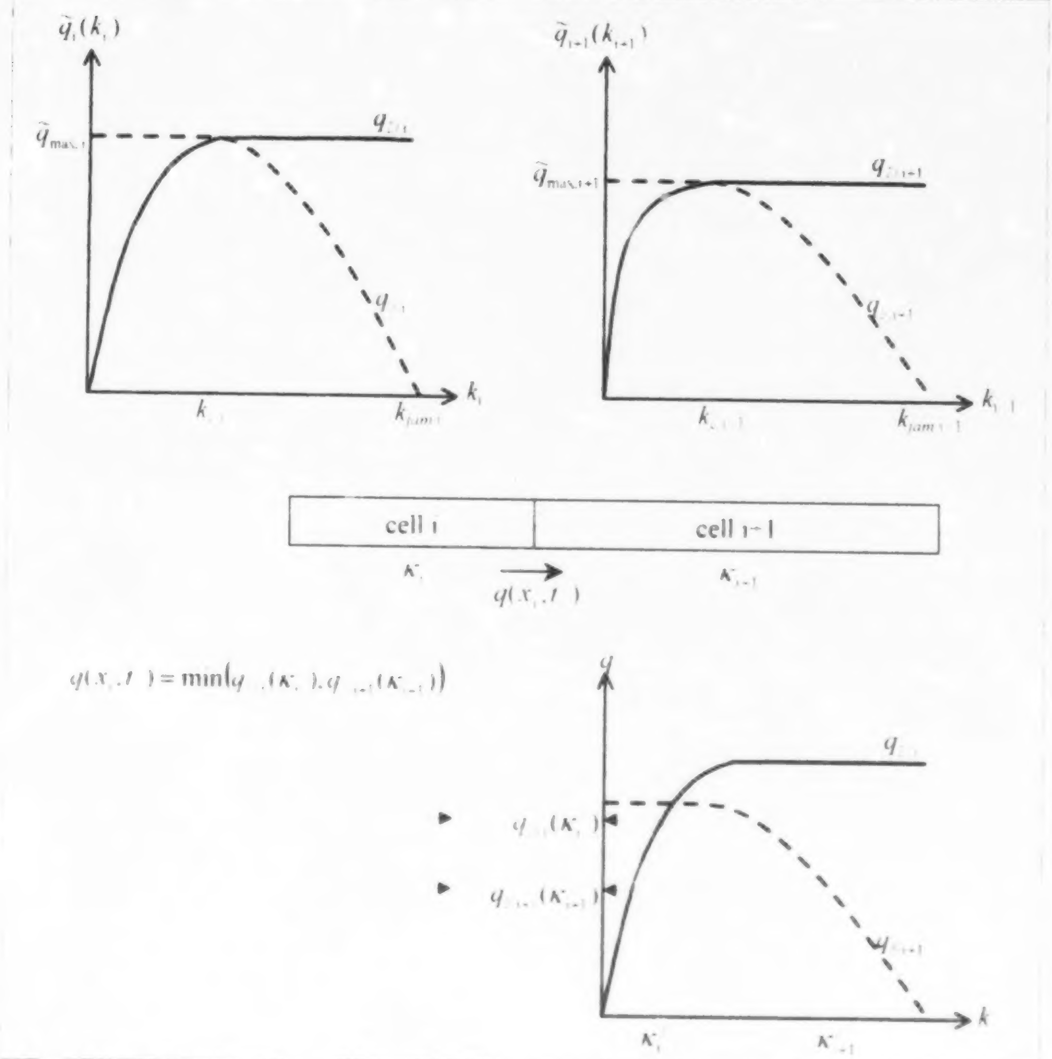
The solution to the generalised Riemann problem is the minimum of the local demand in the upstream cell and the local supply (available space) in the downstream cell Figure 2.16.

[2.27] $$q(x_i, t_j) = \min\left(q_{D,i}(\kappa_i^j), q_{S,i+1}(\kappa_{i+1}^j)\right)$$

[2.27] is equivalent to Table 2.1.

Lebacque's version of the Godunov scheme requires that the flow across a cell boundary, which is the solution to the generalised Riemann problem, remains constant throughout the time step. The density discontinuity at the cell boundary produces a shockwave that may travel either upstream or downstream at speed $u = \dfrac{\Delta q}{\Delta k} \le \dfrac{\partial \tilde{q}(k)}{\partial k}$ as given in [2.13]. If this shockwave reaches an adjacent cell boundary before the next time step, then another generalised Riemann problem presents itself, likely with a different solution. Therefore, Step 1b in Figure 2.14 is valid only if no shockwaves cross a cell boundary during the time step. Hence the discretisation of the roadway must satisfy the constraint

[2.28]
$$\Delta x_i \geq \Delta t \cdot \max_{\substack{k\in[0,k_{jam}] \\ x\in[x_i,x_{i+1}] \\ t\in[t_j,t_{j+1}]}} \frac{\partial \tilde{q}(k,x,t)}{\partial k}\ \forall i .$$



$\tilde{q}_i(k_i)$
$\tilde{q}_{max,i}$
$q_{D,i}$
$q_{S,i}$
$k_{c,i}$
$k_{jam,i}$
k_i
$\tilde{q}_{i+1}(k_{i+1})$
$\tilde{q}_{max,i+1}$
$q_{D,i+1}$
$q_{S,i+1}$
$k_{c,i+1}$
$k_{jam,i+1}$
k_{i+1}
cell i
cell i+1
κ_i
$q(x_i,t_j)$
κ_{i+1}
$q(x_i,t_j) = \min(q_{D,i}(\kappa_i), q_{S,i+1}(\kappa_{i+1}))$
q
$q_{D,i}$
$q_{D,i}(\kappa_i)$
$q_{S,i+1}(\kappa_{i+1})$
$q_{S,i+1}$
k
κ_i
κ_{i+1}

Figure 2.16 Traffic supply and traffic demand curves at a cell boundary

which means that the length of cell i must be at least equal to the maximum wave speed for any possible density at any position within the cell at any time within the time step. If the equilibrium flow-density relationship is homogeneous over the length of the cell and the duration of the time step, then [2.28] reduces to the CFL stability condition [2.22].

2.4 The Cell-transmission Model

The name "cell-transmission model" was coined by Daganzo (1994). Several papers (Daganzo, 1994; 1995a; 1995b) on the development and theory of the cell-transmission model, are summarised in this section. The formulation of the cell-transmission model is given in the first subsection. The next subsection presents an estimate of the error in the approximation of the LWR hydrodynamic model by the cell-transmission model. The last subsection addresses the practical concerns of implanting the cell-transmission model within a network, which may contain multiple origins, multiple destinations, and various intersection configurations.

2.4.1 Formulation of the cell-transmission model

Daganzo's cell-transmission model is similar to Godunov's finite difference approximation to the LWR hydrodynamic model. The cell placement is identical to that of Godunov in Figure 2.12, with a discretisation scheme that is replaced with an equality

[2.29] $$\Delta x = v_f \Delta t .$$

Each cell i is assigned an equilibrium flow-density relationship $\tilde{q}_i(k)$, which implicitly defines the free-speed v_f, speed-at-capacity v_c, capacity $q_{\max}$, critical density k_c, and jam density k_{jam}. This function is used to generate the local demand $\tilde{q}_{D,i}(k)$ and local supply $\tilde{q}_{S,i}(k)$ curves, according to [2.25] and [2.26], respectively. The state of the cell is defined by the occupancy $n_i(t)$, which may be written as n_i^t for the discretised time interval $[t_j, t_{j+1}]$. Assuming uniform density throughout each cell, the density can be estimated as $K_i^t = n_i^t \quad \Delta x$.

Initial occupancies n_i^0 are given. The occupancy is updated after each time-step according to

[2.30] $$n_i^{t+1} = n_i^t + y_{i-1}^t - y_i^t .$$

The number of exiting vehicles is determined by $y_i^t = q(x_i, t_i)\Delta t$, where the flow $q(x_i, t_i)$ is specified by the solution to the Riemann problem in [2.27].

Substitution of $y_i^t = q(x_i, t_i)\Delta t$ and [2.27] into [2.30] gives

[2.31] $$n_i^{t+1} = n_i^t + \min\left(q_{D,i-1}(K_{i-1}^t), q_{S,i}(K_i^t)\right)\Delta t - \min\left(q_{D,i}(K_i^t), q_{S,i+1}(K_{i+1}^t)\right)\Delta t \ .$$

When implementing this model in a simulation environment, it is important to note the dependency of the future density. In uncongested conditions, the cell density depends only on upstream conditions. So, a sequential update of each cell density in reverse order using [2.31] would yield the correct solution. In congested conditions, the cell density depends only on downstream conditions. So, a sequential update of each cell density, starting with the first cell, would yield the correct solution. However, a cell with congestion downstream and uncongested conditions upstream poses a problem for the order of sequential calculations.

A more effective solution method is to first calculate the exit flows from each cell, then update each cell density. Though this method requires storage of each cell exit flow, it enables parallel implementation since the cells may be considered in any order.

2.4.2 Error quantification

The following discussion compares the cell transmission model with the LWR hydrodynamic model. Firstly, the error in the cell-transmission is quantified along characteristics. Next, the representation of shockwaves, which interrupt the characteristics, is discussed.

2.4.2.1 Error along a characteristic

The discrete lattice points (x_i, t_j) are located at the end of each cell at the beginning of each time step. The true LWR solution $k(x,t)$ is approximated by the continuous function $K(x,t)$, which passes through the lattice points with value $K_i^j = K(x_i, t_j)$. The initial density is given by

$$K(x,0) = \int_{x-\Delta x}^{x} k(x,0)\,\mathrm{d}x \,,$$

which implies that the initial occupancy $n_i^0 = K(x_i, t_0)\Delta x$ is equal to the average density in cell i at time $t_0 = 0$. Obviously, as the cell length Δx approaches zero, the approximated initial density $K(x,0)$ approaches $k(x,0)$.

The densities are updated according to [2.30], which can be written as

[2.32] $$K_i^{j+1} = K_i^j + \frac{y_{i-1}^j - y_i^j}{\Delta x}\,.$$

Note that [2.32] is identical to Step 1 of Lebacque's modified Godunov scheme in Figure 2.14. Similarly, [2.31] can be written as

[2.33] $$K_i^{j+1} = K_i^j + \frac{\min\left(q_{D,i-1}(K_{i-1}^j), q_{S,i}(K_i^j)\right)\Delta t - \min\left(q_{D,i}(K_i^j), q_{S,i+1}(K_{i+1}^j)\right)\Delta t}{\Delta x}\,.$$

For a homogeneous road with a smooth, concave equilibrium flow-density relationship, assume for the moment that the road is either in the uncongested regime, the congested regime, or at capacity. Under uncongested conditions $K < k_c$, the first argument of each min() operator in [2.33] is the active constraint. The second argument is active for congested conditions $K > k_c$.

If the density is equal to the critical density, then the arguments are equal; hence both arguments are active. So, [2.33] can be split into two cases

$$K_i^{j+1} = \begin{cases} K_i^j + \left[\tilde{q}(K_{i-1}^j) - \tilde{q}(K_i^j)\right]\Delta t/\Delta x, & \text{if } K \le k_c \\ K_i^j + \left[\tilde{q}(K_i^j) - \tilde{q}(K_{i+1}^j)\right]\Delta t/\Delta x, & \text{if } K \ge k_c \end{cases}$$

which can be expressed more compactly as

[2.34] $$K_i^{j+1} = K_i^j \pm \left[\tilde{q}(K_i^j) - \tilde{q}(K_{i\pm1}^j)\right]\Delta t/\Delta x,$$

where the ± operator implies addition if $K \ge k_c$ or subtraction if $K \le k_c$.

During this time-step $[t_j, t_{j+1}]$, the characteristic which reaches (x_i, t_{j+1}) must have started the time-step at $(x_i - w\Delta t, t_j)$. By subtracting the value of the density $K(x_i - w\Delta t, t_j)$ from both sides of [2.34], an expression is derived for the error in density introduced by the FDE approximation to the LWR hydrodynamic model during a single time-step

$$K_i^{j+1} - K(x_i - w\Delta t, t_j) = \left[K_i^j - K(x_i - w\Delta t, t_j)\right] \pm \left[\tilde{q}(K_i^j) - \tilde{q}(K_{i\pm1}^j)\right]\Delta t/\Delta x.$$

Daganzo (1995b) suggested that each bracketed term can be expressed to a second order accuracy in Δt and Δx through a power series expansion about $(x_i - w\Delta t, t_j)$, assuming that $K(x,t)$ is smooth along the characteristic. The result is presented as follows

$$K_i^{j+1} - K(x_i - w\Delta t, t_j) \cong -\tfrac{1}{2}K_{xx}[\Delta t^2\tilde{q}_k^{\,2} \pm \tilde{q}_k\Delta t\Delta x] - [\tilde{q}_k\Delta t/\Delta x \pm \tfrac{1}{2}]\tilde{q}_{kk}(K_x)^2\Delta t\Delta x,$$

where subscripts indicate partial derivatives. This simplifies to

[2.35] $$K_i^{j+1} - K(x_i - w\Delta t, t_j) \cong \tfrac{1}{2}p(1-p)K_{xx}\Delta x^2 \pm [p - \tfrac{1}{2}]\tilde{q}_{kk}(K_x)^2\Delta t\Delta x.$$

where $p = |\tilde{q}_k| \Delta t / \Delta x$ is the fraction of a cell which is crossed by a characteristic during one time-step. The cumulative error along the characteristic can be estimated by dividing [2.35] by Δt and integrating the expression from 0 to the current time. The resulting expression should not be used to evaluate the cumulative error along a characteristic emanating from a fan, since the density at the fan is not smooth. Nevertheless, Daganzo (1995b) informally justified the FDE approximation to the LWR solution for characteristics emanating from a fan using qualitative arguments.

Daganzo (1995b) accounted for the error terms on the right hand side of [2.35]. The first term is deemed a "diffusion effect," which tends to smooth density with increasing time. The second term is called a "kinematic effect." This FDE approximation will be stable only if the diffusion coefficient is positive, i.e. if $p \le 1$ for all $\tilde{q}_k$. This constraint is satisfied by the CFL stability condition [2.22]. The error due to the diffusion effect is maximised for $p = \frac{1}{2}$, but it disappears when $p = 0$ or $p = 1$.

2.4.2.2 Shockwave representation

The cell-transmission model approximates the density profile of a shockwave with arbitrary precision, depending on the discretisation. In the LWR hydrodynamic model, a shockwave is a moving (or stationary) density discontinuity along the space-time plane with speed u. Shockwaves may occur in the LWR model in two ways: (1) as a result of a density discontinuity in the initial density profile $k(x,0)$ with a positive jump in density, or (2) due to a convergence of characteristics which occurs if $(k_0)_x = \dfrac{\partial k(x,0)}{\partial x} > 0$.

In the cell-transmission model any convergence of characteristics is countered by the diffusion effect. Eventually, the convergence of characteristics will be balanced by an equivalent diffusion effect, thereby preventing the formation of a sharp (discontinuous) shock. Hence, shocks span multiple cells in the cell-transmission model. If, as in (1) above, the initial density contains a discontinuity, then the diffusion effect will dissipate the shock, creating a steep density profile. Eventually, the converging characteristics at the steep density profile are balanced by the diffusion effect. A shock which has reached this balanced state is called a *stable* shock.

Though shocks never spread in the LWR model, empirical evidence shows that shockwaves do dissipate in reality. The lack of dissipation in the LWR model is due to the assumption that vehicles can accelerate and decelerate at infinite rates. Though the cell-transmission model does dissipate shockwaves due to the diffusion effect, this dissipation is a result of numerical error, which can be reduced by using a finer discretisation. Hence, the mechanism which causes the actual dissipation of shockwaves is not explicitly modelled by the cell-transmission model.

The distribution of the shock across the cells can be derived (Daganzo, 1995b). Consider the case of a homogeneous roadway with concave $\tilde{q}(k)$, and initial density $k_{0x} > 0$, $k(x,0) < k_c$, which implies a forward-moving shockwave, i.e. $u > 0$. The density is updated according to [2.34]

$$K_i^{t+1} = K_i^t + \left[\tilde{q}(K_{i-1}^t) - \tilde{q}(K_i^t)\right]\Delta t / \Delta x .$$

Subtracting $K(x - u\Delta t, t_j)$ from both sides and dividing by Δt gives

$$\frac{K_i^{t+1} - K(x - u\Delta t, t_j)}{\Delta t} = \frac{K_i^t - K(x - u\Delta t, t_j)}{\Delta t} + \frac{\left[\tilde{q}(K_{i-1}^t) - \tilde{q}(K_i^t)\right]\Delta t}{\Delta x} .$$

In the limit as Δt approaches zero, the left-hand side is the partial derivative of density with respect to time along the path of the shockwave. The expression of the bracketed terms on the right-hand side as a second order power series expansion about $(x - u\Delta t, t_j)$ yields

$$K_t + uK_x = [u - \tilde{q}(K)]K_x + \frac{\Delta t}{2}\left[u^2 - \left(2u - \Delta x/\Delta t\right)\tilde{q}_k(K)\right]K_{xx} - \frac{\Delta t}{2}\left[2u - \Delta x/\Delta t\right]\tilde{q}_{kk}(K)K_x^2 .$$

Recall that subscripts $_x$, $_t$, and $_k$ indicate partial derivatives. When a shock becomes stable, the variation in density along the shock is zero $K_t + uK_x = 0$, so the above equation simplifies to an ordinary differential equation

$$0 = [uK - \tilde{q}_k(K)]_x + \frac{\Delta t}{2}\left[u^2K - \left(2u - \Delta x/\Delta t\right)\tilde{q}(K)\right]_{xx} .$$

Integration by x with the boundary condition that $K_x = 0$ when $K = k_r$, where k_r is the density downstream (right) of the shock, gives

$$1 = \frac{\Delta t}{2}\frac{\left[u^2 - \left(2u - \Delta x/\Delta t\right)\tilde{q}_k(K)\right]K_x}{u(k_r - K) - (\tilde{q}(k_r) - \tilde{q}(K))}$$

Given two densities k_a and k_b, such that $k_l < k_a < k_b < k_r$, one may obtain the length of roadway, measured in cells, which spans the density transition from k_a to k_b by integrating the above expression between x_a and x_b, then dividing by Δx

[2.36] $$\frac{x_b - x_a}{\Delta x} = \frac{\Delta t}{2\Delta x}\int_{k_a}^{k_b}\frac{\left[u^2 - \left(2u - \Delta x/\Delta t\right)\tilde{q}_k(K)\right]}{u(k_r - K) - (\tilde{q}(k_r) - \tilde{q}(K))}\,\mathrm{d}K .$$

If the initial density had been specified such that $k(x,0) > k_c$, which implies a backward-moving shockwave, then a similar derivation would yield

[2.37] $$\frac{x_b - x_a}{\Delta x} = \frac{\Delta t}{2\Delta x}\int_{k_a}^{k_b}\frac{\left[u^2 - \left(2u + \Delta x/\Delta t\right)\tilde{q}_k(K)\right]}{u(k_l - K) - (\tilde{q}(k_l) - \tilde{q}(K))}\,\mathrm{d}K .$$

If the densities are chosen such that $k_a = k_l$ or $k_b = k_r$, then [2.36] and [2.37] give an infinite number of cells. Daganzo (1995b) reported that for $u > 0$, 99% of the shock covered four cells or fewer, but more cells were required to represent backward-moving shockwaves. An infinite number of cells were required to capture a shockwave moving at the maximum backward speed.

Notice that the number of cells spanning densities k_a and k_b is a function of the discretisation in the form of the ratio $\Delta x/\Delta t$. Examination of the numerator in the integrand in [2.36] reveals that the smoothing of the shock increases with $\Delta x/\Delta t$, provided that $u^2/\tilde{q}_k(K) - 2u + \Delta x/\Delta t \geq 0$. This condition is easily satisfied since $u \cong \tilde{q}_k(K)$ for $k_l < k_r < k_c$, which reduces the condition to $\Delta x/\Delta t \geq u$. (Recall that the CFL stability condition [2.22] restricts the discretisation $\Delta x/\Delta t \geq v_f$, and that $v_f > u$.) A similar analysis of [2.37] supports the finding that increases in $\Delta x/\Delta t$ tend to smooth the shock. Thus, shocks may be made arbitrarily sharp using finer discretisations in the same ratio $\Delta x/\Delta t$. In the limit as $\Delta x, \Delta t \rightarrow 0$, the shock becomes a density discontinuity as in the LWR model.

The cell length Δx can be defined as a function of Δt such that the smoothing of shocks is minimised. The number of cells spanning densities k_a and k_b can be minimised by minimising $\Delta x/\Delta t$. Recall that the ratio $\Delta x/\Delta t$ is constrained below by the CFL stability condition [2.22]. So, to reduce shock smoothing the cell length must be $\Delta x = v_f \Delta t$, as noted in [2.29].

This discretisation scheme has a major drawback in that the actual arc length must be approximated by an integer multiple of the cell length, which is a function of the time-step Δt, a network-wide parameter. This can introduce considerable modelling inaccuracy for coarse discretisations, necessitating a judicious choice of the time step Δt.

Now, consider the case where $k_l < k_c < k_r$. If $u > 0$, then the transition from $K < k_c$ to $K = k_r$ is abrupt on the downstream side of the shock, while the transition from $K = k_a$ to $K = k_c$ on the upstream side of the shock is governed by [2.36]. Similarly, if $u < 0$, then the transition from $K = k_l$ to $K > k_c$ is abrupt on the upstream side of the shock, while the transition from $K = k_c$ to $K = k_b$ on the downstream side of the shock is governed by [2.37].

The abrupt and gradual density transitions on each side of a shock can be intuitively understood with the aid of influence diagrams. Figure 2.17 and Figure 2.18 contain the flow-density diagram, characteristics, and influence diagram for four shockwaves. The lattice points on the influence diagram represent the downstream end of each cell for the beginning of each time-step. Lattice points are separated by Δx on the horizontal axis and by Δt on the vertical axis. The positive diagonal between lattice points has slope equal to the free speed $v_f = \Delta x / \Delta t$. Arrows represent a causal influence on the density, according to [2.34]. Where two dashed arrows cross in Figure 2.18, only one of the pair is actually a valid influence. The valid arrow is determined by the actual density values.

Each of the four *stable* shockwaves in Figure 2.17 and Figure 2.18 pass between lattice points (x_i, t_j) and (x_{i-1}, t_j). In each influence diagram, the lattice point (x_i, t_j) is circled, and the lattice point (x_{i-1}, t_j) is boxed. A time $t_{j-3} < t_j$ was chosen arbitrarily to demonstrate how influence zones determine the type of density transition upstream and downstream of the shock.

The influence zone contains all lattice points which may influence the density at a given future lattice point. The influence zone for lattice point (x_i, t_j) is displayed as a striped cone extending backwards in time. The influence zone for lattice point (x_{i-1}, t_j) is displayed as a tinted cone.

The lattice points at time t_{j-3}, which influence lattice points (x_i, t_j) and (x_{i-1}, t_j) have been circled and boxed, respectively, in each influence diagram.

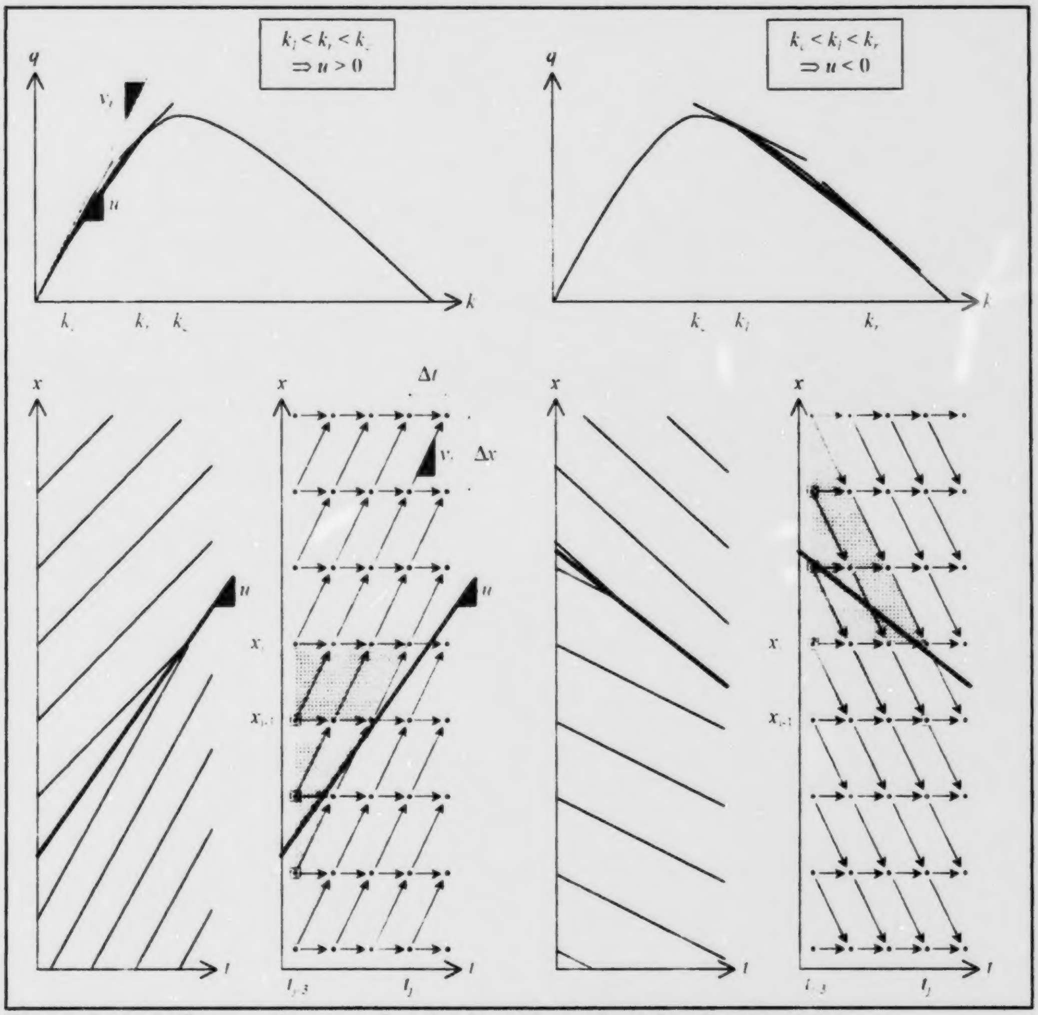

$k_l < k_r < k_c$
$\Rightarrow u > 0$
q
v_l
u
k
k_l
k_r
k_c
$k_c < k_l < k_r$
$\Rightarrow u < 0$
q
k
k_c
k_l
k_r
x
u
t
x
Δt
v_l
Δx
x_i
x_{i-1}
u
t
t_{j-3}
t_j
x
t
x
x_i
x_{i-1}
t
t_{j-3}
t_j

Figure 2.17 Influence diagrams for shocks with gradual density transitions

If the influence zone of a point spans the shockwave path, then that point is influenced by some lattice points with density k_l and others with density k_r. This mixed influence produces the gradual density transition near the path of the shockwave.

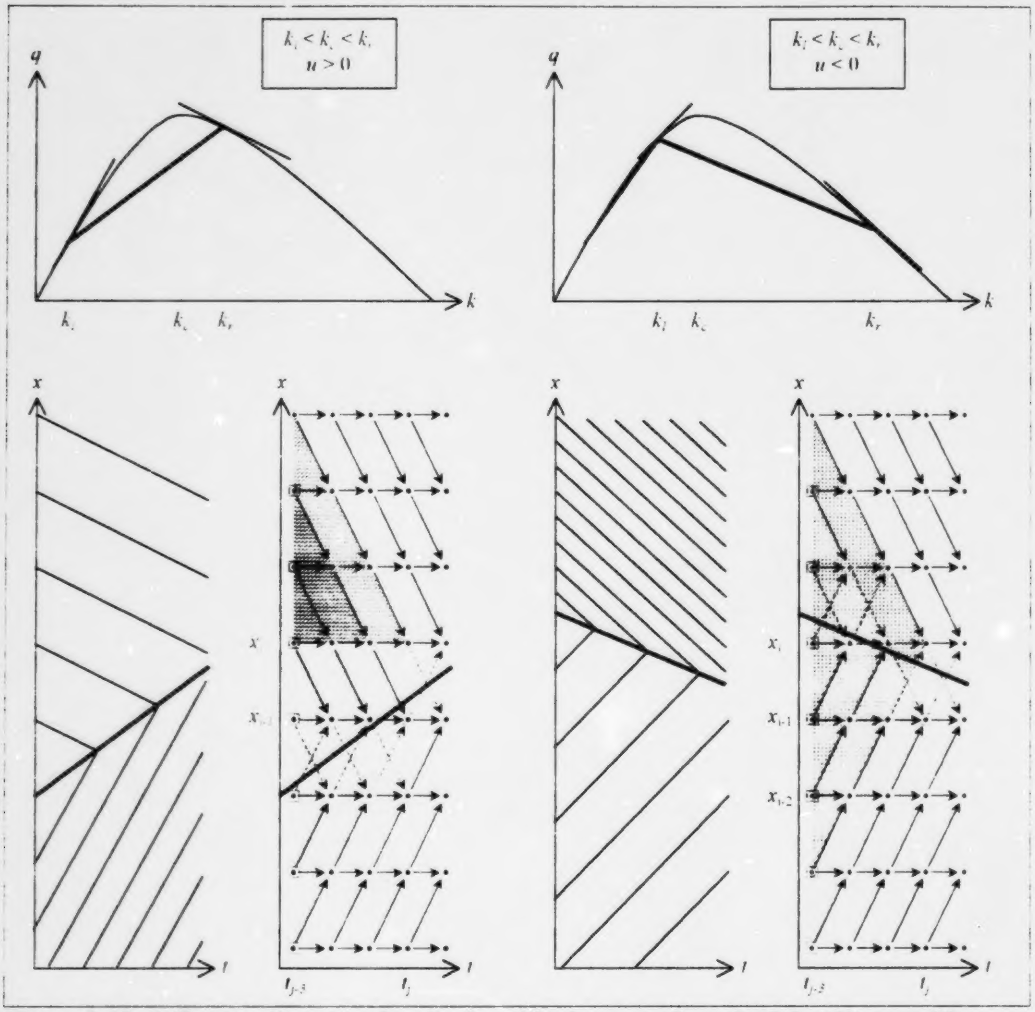

$k_l < k_c < k_r$
$u > 0$
$k_l < k_c < k_r$
$u < 0$
q
k
k_l
k_c
k_r
q
k
k_l
k_c
k_r
x
t
x
x_i
x_{i-1}
t
t_{j-3}
t_j
x
t
x
x_i
x_{i-1}
x_{i-2}
t
t_{j-3}
t_j

Figure 2.18 Influence diagrams for shocks with abrupt and gradual density transitions

The influence zones in Figure 2.17 each span the shockwave path, implying that the lattice points (x_i, t_j) and (x_{i-1}, t_j) are influenced by some lattice points with density k_l and others with density k_r. This mixed influence produces the gradual density transition near the path of the shockwave. The gradual density transitions on both sides of the shock is consistent with [2.36] and [2.37]. For every lattice point arbitrarily far from the shockwave path in Figure 2.17, the influence zone spans the shockwave path if extended far enough in the past. This is true because $-v_f < u < v_f$.

In Figure 2.18 the first influence diagram has a tinted influence zone that spans the shockwave path, while the striped influence zone of lattice point (x_i, t_j) lies solely downstream of the shock. This demonstrates why the upstream density transition is gradual, and the downstream density transition occurs abruptly across a single cell. In the second influence diagram of Figure 2.18 both influence zones span the shockwave path, however the unmarked influence zone of upstream lattice point (x_{i-2}, t_j) lies solely on the upstream side of the shockwave path. Therefore, the downstream density transition is gradual, whereas the upstream density transition occurs abruptly across two cells. As the slope of the shockwave decreases, abrupt density transitions occur across more cells.

Daganzo (1995b) argues that these findings for the error in density along a characteristic and the representation of a shockwave can be extended in several ways. The assumption of a smooth equilibrium flow-density curve can be relaxed, as can the assumption concerning a homogeneous roadway, provided that (1) the free-speed v_f remains bounded from above so that the CFL stability condition is maintained, and (2) the jam density k_{jam} is not suddenly reduced below the current density.

It is also important to address the representation of shockwaves for the case of piecewise-linear equilibrium flow-density relationships. As mentioned previously, some shocks may not respect the entropy condition or driver's ride impulse in this case. For these shocks, the diagonal influence lines point away from the shock, except for the two lattice points nearest the shock. So, the influence zone of any point arbitrarily far from the shock, either upstream or downstream, will eventually span the shockwave if extended far enough into the past. Therefore, the density transition is gradual on both sides of shocks which do not satisfy the entropy condition.

One exception to the above rule is shocks which travel at the free-speed. These fast-moving shocks occur if the density discontinuity lies entirely within the leftmost linear section of the piecewise-linear equilibrium flow-density relationships. These shocks have abrupt density transitions across a single cell that spans the shock, because the influence zone of any point, which is at least one cell length away from the shock, will never span the shock.

2.4.3 Approximation of the solution of the LWR model

The findings of the previous sections can be used to justify that the cell-transmission model approximates the physically relevant solution of the LWR hydrodynamic model. If multiple solutions exist, which is the case if the equilibrium flow-density relationships is non-linear and the density has discontinuity with a negative jump, then the physically relevant solution is the entropy solution.

It was shown previously that the error in the density along a characteristic can be made arbitrarily small by making the discretisation finer. Provided that the approximated characteristics form the physically relevant solution of the LWR model, then the cell-transmission model yields an approximation to that solution. Since characteristics begin and end at density discontinuities, it is sufficient to demonstrate that the approximated characteristics are consistent with the physically relevant solution in their emergence from density discontinuities.

It was found in the previous section that initial conditions, which cause a shockwave to form in the LWR model, also produce a stable shockwave in the cell-transmission model. Though stable shockwaves appear as steep, rising density profiles rather than density discontinuities, the steepness of the density profile is conserved with time and the profile can be made arbitrarily steep by making the discretisation finer.

The same logic can be used to show that the cell-transmission model produces a fan whenever the initial conditions specify a negative jump in density at a discontinuity for a non-linear equilibrium flow-density relationship (Daganzo, 1995b). The smoothing of the density discontinuity, which occurs at a fan, is enhanced by the diffusion effect. The diffusion effect can be decreased by making the discretisation finer.

2.4.4 Adaptation to traffic networks

Thus far, the cell transmission model has considered only flow on linear roadways with no beginning or end. For this model to be practical it must be adapted to networks. Daganzo (1995a) addressed several concerns in this regard. The following section describes the transfer of flow across nodes. Next, the creation of vehicles at origins and their absorption at destinations is considered. The introduction of multiple routes to one or more destinations necessitates a means for tracking the flow by route. The last section discusses the structure within cells to manage the multi-commodity flow.

2.4.4.1 Simple intersection models

Daganzo (1995a) extended the concepts of local supply and local demand curves to formulate node models to account for freeway intersections. Intersections are modelled using pointers from the upstream arcs to the downstream arcs. All connections between arcs must be either ordinary (one upstream arc empties into one downstream arc), merge (multiple upstream arcs empty into one downstream arc), or diverge (one upstream arc empties into multiple downstream arcs). These three simple intersections are shown in Figure 2.19, where the pointers are labelled the vehicle movements across the intersection y_{ij} from cell i into cell j. Though the following discussion considers interactions with up to two upstream or downstream arcs, the same node models can be generalised to apply to merge intersections with any number of upstream arcs and to diverge intersections with any number of downstream arcs.

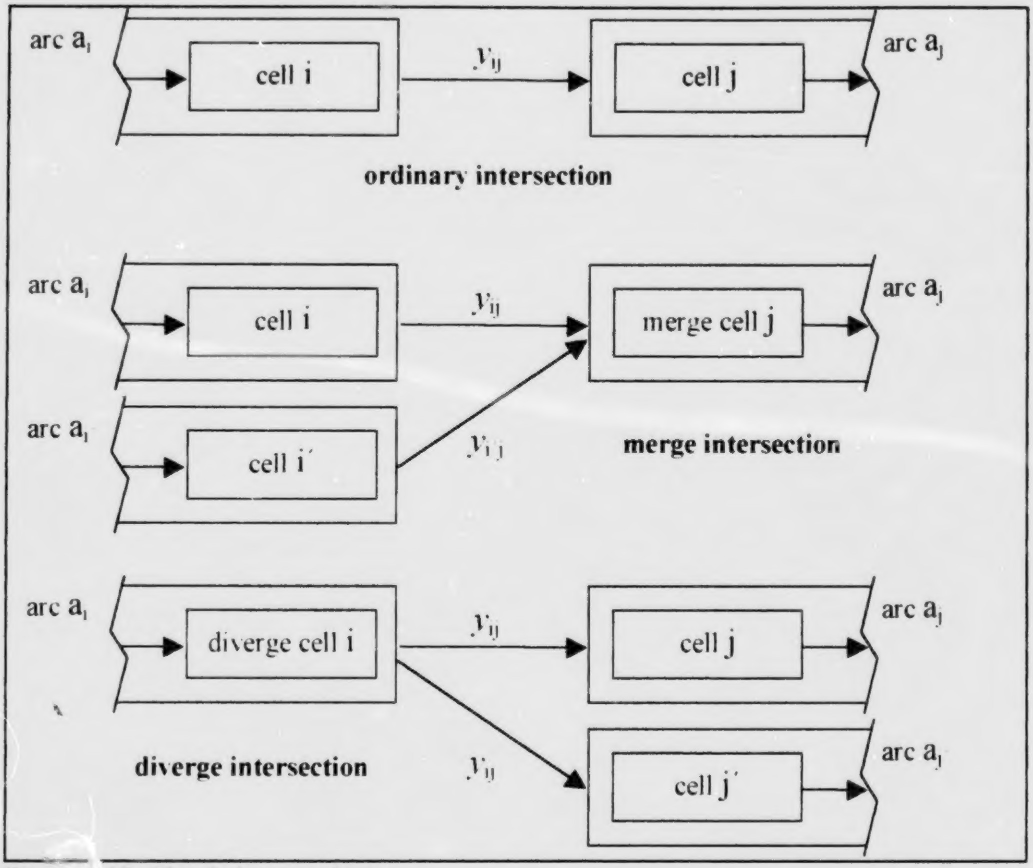

arc a_i
cell i
y_{ij}
cell j
arc a_j
ordinary intersection
arc a_i
cell i
y_{ij}
merge cell j
arc a_j
arc a_i
cell i′
$y_{i'j}$
merge intersection
arc a_i
diverge cell i
y_{ij}
cell j
arc a_j
diverge intersection
$y_{ij'}$
cell j′
arc a_j

Figure 2.19 Simple intersections

The ordinary intersection is treated in the same way as two cells within an arc. The exit flow from cell i into cell j during time step $[t_j, t_{j+1}]$ is calculated according to [2.27]

$$\frac{y_{ij}^{t}}{\Delta t} = \min\left(\tilde{q}_{D,i}, \tilde{q}_{S,j}\right)$$

where $\tilde{q}_{D,i}$ is shorthand for $\tilde{q}_{D,i}(K(x_i, t_j))$. Cell occupancy is updated according to [2.30], restated as $n_j^{t+1} = n_j^t + y_i^t - y_j^t$.

The occupancy of merge cell j is updated according to

$$n_j^{t+1} = n_j^t + y_{ij}^t + y_{i'j}^t - y_j^t .$$

The vehicle movements at a merge y_{ij}^t and $y_{i'j}^t$ must satisfy the following constraints

$$[2.38] \qquad \begin{array}{ll} y_{ij} \le \tilde{q}_{D,i}\Delta t,\ y_{i'j} \le \tilde{q}_{D,i'}\Delta t & \text{local demand constraints} \\ y_{ij} + y_{i'j} \le \tilde{q}_{S,j}\Delta t & \text{local supply constraint} \\ y_{ij} \ge 0, y_{i'j} \ge 0 & \text{non - negativity constraints} \end{array}$$

which defines a feasible region. (Note that the superscript denoting the time step is dropped for convenience.) The challenge of node models is to determine which point in the feasible region most closely matches realistic intersection flows at any instant.

Drivers tend to behave selfishly, which produces a user-optimal solution for the dynamic network loading problem. Intersection design, including the traffic control scheme, aims to achieve a safe interaction among drivers while forcing them to follow rules which lead to a system-optimal solution. For some simple networks, such as merges and diverges, the user-optimal and system-optimal solutions coincide with the max flow solution. Still, the max flow solution in the feasible region defined by [2.38] is not necessarily unique. Though extreme points may qualify as optimal solutions, they are unrealistic because they heavily favour the flow of one stream over another. A more realistic solution allows several streams to share the available discharge capacity and downstream local supply. The node model can select such an optimal solution by distributing the available discharge capacity and downstream local supply according to distribution factors α_{ij} and β_{ij}, respectively. The distribution factors may be weighted according to appropriate criteria, for example partial demands or discharge capacities, and they should respect $\sum_{ij} \alpha_{ij} = 1$ and $\sum_{ij} \beta_{ij} = 1$.

For merge intersections, Lebacque (1996) suggested combining [2.27] with partial flows entering the merge cell j from cell i and cell i' given by

[2.39]
$$
\begin{aligned}
\frac{y_{ij}}{\Delta t} &= \min(\tilde{q}_{D,i}, \beta_{ij}\tilde{q}_{S,j}) \\
\frac{y_{i'j}}{\Delta t} &= \min(\tilde{q}_{D,i'}, \beta_{i'j}\tilde{q}_{S,j})
\end{aligned}
$$

where the local supply of the merge cell j is shared between the upstream cells i and i' according to local supply distribution factors β_{ij} and $\beta_{i'j}$, such that $\beta_{ij} + \beta_{i'j} = 1$, to yield a unique solution to the max flow problem

$$
\begin{array}{lll}
\max y_{ij} + y_{i'j} & & \\
\text{subject to } y_{ij} \le \tilde{q}_{D,i}\Delta t,\ y_{i'j} \le \tilde{q}_{D,i'}\Delta t & & \text{local demand constraints} \\
\qquad y_{ij} \le \beta_{ij}\tilde{q}_{S,j}\Delta t,\ y_{i'j} \le \beta_{i'j}\tilde{q}_{S,j}\Delta t & & \text{local supply constraints} \\
\qquad y_{ij} \ge 0,\ y_{i'j} \ge 0 & & \text{non - negativity constraints}
\end{array}
$$

If the first term of the min() operator is active for both partial flows in [2.39], then the local demand of upstream cells i and i' is exhausted. If the second term is active for both partial flows, then the local supply of merge cell j is exhausted. If the first term is active for one flow, while the second term is active for the other, then some portion of the local supply of merge cell j is unused even though there is residual local demand on one of the upstream arcs. This result is inconsistent with real traffic merges, in which driver behaviour is consistent with a user-optimal paradigm.

The corresponding graph is shown in Figure 2.20. Each stream at the merge corresponds to a path in the graph.

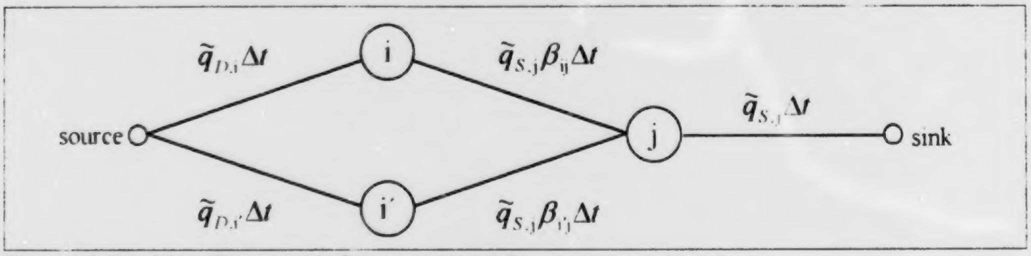

$\tilde{q}_{D,i}\Delta t$
i
$\tilde{q}_{S,j}\beta_{ij}\Delta t$
$\tilde{q}_{S,j}\Delta t$
source
j
sink
$\tilde{q}_{D,i'}\Delta t$
i′
$\tilde{q}_{S,j}\beta_{i'j}\Delta t$

Figure 2.20 Max flow graph for a merge node

Lebacque suggests weighting the local supply distribution factors by the discharge capacity of each upstream arc. This produces an unused residual local supply described above. Weighting the distribution factors by the partial demands of each stream alleviates the unused residual local supply, but heavily favours a congested stream over an uncongested stream.

The naive behaviour described above can be ameliorated with a more complex merge model proposed by Daganzo (1995a), in which the partial flows are given by

$$\frac{y_{ij}}{\Delta t} = \begin{cases} \tilde{q}_{D,i} & \text{if } \tilde{q}_{D,i} \le \beta_{ij}\tilde{q}_{S,j} \\ \beta_{ij}\tilde{q}_{S,j} & \text{if } \tilde{q}_{D,i} \ge \beta_{ij}\tilde{q}_{S,j} \text{ and } \tilde{q}_{D,i'} \ge \beta_{i'j}\tilde{q}_{S,j} \\ \tilde{q}_{S,j} - \tilde{q}_{D,i'} & \text{if } \tilde{q}_{D,i} \ge \beta_{ij}\tilde{q}_{S,j} \text{ and } \tilde{q}_{D,i'} \le \beta_{i'j}\tilde{q}_{S,j} \end{cases}$$

$$\frac{y_{i'j}}{\Delta t} = \begin{cases} \tilde{q}_{D,i'} & \text{if } \tilde{q}_{D,i'} \le \beta_{i'j}\tilde{q}_{S,j} \\ \beta_{i'j}\tilde{q}_{S,j} & \text{if } \tilde{q}_{D,i'} \ge \beta_{i'j}\tilde{q}_{S,j} \text{ and } \tilde{q}_{D,i} \ge \beta_{ij}\tilde{q}_{S,j} \\ \tilde{q}_{S,j} - \tilde{q}_{D,i} & \text{if } \tilde{q}_{D,i'} \ge \beta_{i'j}\tilde{q}_{S,j} \text{ and } \tilde{q}_{D,i} \le \beta_{ij}\tilde{q}_{S,j} \end{cases} \quad [2.40]$$

In the first case of [2.40], local supply exceeds local demand and both cells i and i' are in the uncongested regime. In the second case, local demand exceeds local supply and both cells i and i' are in the congested regime. In the third case, one cell is congested and the other is uncongested. The partial flows are constrained by the local supply distribution factors if and only if the local demand exceeds the proportion of the local supply on both approaches to the merge, as in the second case. Otherwise, any residual local supply in merge cell j is used by the merge approach with surplus local demand.

This merge model can also be represented with a max flow graph as in Figure 2.20 by relaxing the capacity of the second arc if the capacity of the first arc on the other paths is not exhausted. This is equivalent to the simplest case of an iterative algorithm which reallocates the residual local supply to other streams.

Diverge cells may also be modelled in a variety of ways. Two such models are presented here. In both models the density of diverge cell i is updated according to

$$n_i^{t+1} = n_i^t + y_{i-1}^t - y_{ij}^t - y_{ij'}^t .$$

In the first model, which is described by Lebacque (1996), the number of exiting vehicles $y_i = y_{ij} + y_{ij'}$ from diverge cell i is determined by

$$\frac{y_{ij}}{\Delta t} = \min\left(\alpha_{ij}\tilde{q}_{D,i}, \tilde{q}_{S,j}\right) \qquad [2.41]$$
$$\frac{y_{ij'}}{\Delta t} = \min\left(\alpha_{ij'}\tilde{q}_{D,i}, \tilde{q}_{S,j'}\right)$$

where the discharge capacity distribution factors are weighted by the partial demands of the diverge cell i, i.e. $\alpha_{ij} = \dfrac{n_{ij}}{n_{ij} + n_{ij'}}$ and $\alpha_{ij'} = \dfrac{n_{ij'}}{n_{ij} + n_{ij'}}$. This model is equivalent to solving the max flow problem

$$\begin{array}{ll}
\max y_i = y_{ij} + y_{ij'} & \\
\text{subject to } y_{ij} \le \alpha_{ij}\tilde{q}_{D,i}\Delta t,\ y_{ij'} \le \alpha_{ij'}\tilde{q}_{D,i}\Delta t & \text{local demand constraints} \\
y_{ij} \le \tilde{q}_{S,j}\Delta t,\ y_{ij'} \le \tilde{q}_{S,j'}\Delta t & \text{local supply constraints} \\
y_{ij} \ge 0,\ y_{ij'} \ge 0 & \text{non - negativity constraints}
\end{array}$$

with the corresponding graph shown in Figure 2.21. This model assumes that each partial demand can reach the desired downstream arc regardless of the local supply of the other downstream arcs. In reality, if the local supply of a downstream arc is insufficient to meet the upstream partial demand, then the remainder of the partial demand tends to block the upstream arc exit. This interference with other partial demands is ignored in this diverge model.

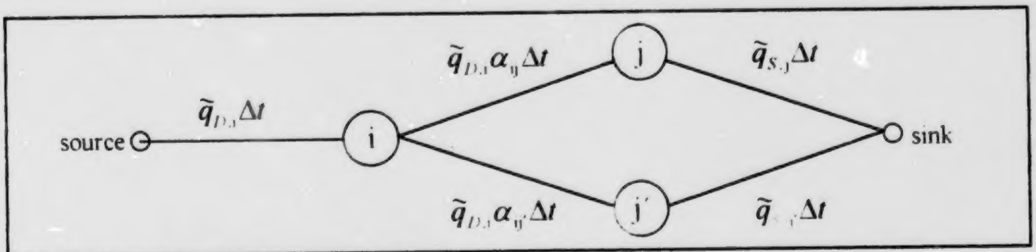

$\tilde{q}_{D,i}\alpha_{ij}\Delta t$
j
$\tilde{q}_{S,j}\Delta t$
$\tilde{q}_{D,i}\Delta t$
source
i
sink
$\tilde{q}_{D,i}\alpha_{ij'}\Delta t$
j′
$\tilde{q}_{S,j'}\Delta t$

Figure 2.21 Max flow graph for a diverge node

An alternative diverge model, which was described by Daganzo (1995a), has exit flow from the diverge cell i given by

[2.42]
$$\frac{y_i}{\Delta t} = \min\left(\tilde{q}_{D,i}, \frac{\tilde{q}_{S,j}}{\alpha_{ij}}, \frac{\tilde{q}_{S,j'}}{\alpha_{ij'}}\right).$$

where the vehicle movements are $y_{ij} = \alpha_{ij}.y_i$ and $y_{ij'} = \alpha_{ij'}.y_i$. [2.41] is equivalent to the solution of the max flow problem

$$\begin{array}{rll} \max & y_i = y_{ij} + y_{ij'} & \\ \text{subject to} & y_{ij} \le \tilde{q}_{D,i}\Delta t & \text{local demand constraint} \\ & y_{ij} \le \tilde{q}_{S,j}\Delta t,\ y_{ij'} \le \tilde{q}_{S,j'}\Delta t & \text{local supply constraints} \\ & y_{ij} = \alpha_{ij}.y_i,\ y_{ij'} = \alpha_{ij'}.y_i & \text{proportionality constraints} \\ & y_{ij} \ge 0,\ y_{ij'} \ge 0 & \text{non - negativity constraints} \end{array}$$

with the corresponding graph shown in Figure 2.22. The capacities of the first arcs in Figure 2.22 are obtained by substitution of the proportionality constraints into the local demand constraint. The capacity of the next arcs is obtained by substitution of the proportionality constraints into the local supply constraints. The capacity of the last arcs is the local supply constraints. The non-negativity constraints are respected implicitly because the two paths from the source node to the sink node are independent.

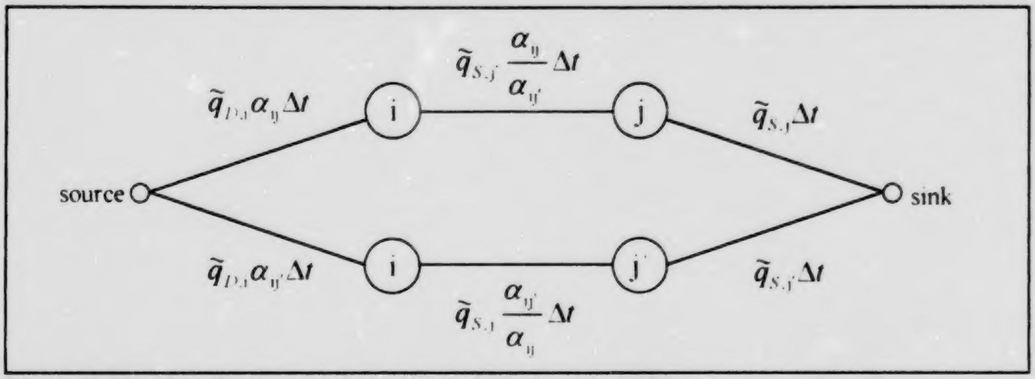

$\tilde{q}_{S,j'}\frac{\alpha_{ij}}{\alpha_{ij'}}\Delta t$
$\tilde{q}_{D,i}\alpha_{ij}\Delta t$
i
j
$\tilde{q}_{S,j}\Delta t$
source
sink
$\tilde{q}_{D,i}\alpha_{ij'}\Delta t$
i
j'
$\tilde{q}_{S,j'}\Delta t$
$\tilde{q}_{S,j}\frac{\alpha_{ij'}}{\alpha_{ij}}\Delta t$

Figure 2.22 Max flow graph for a single-lane diverge node

The implicit assumption of this diverge model is that the flow leaving diverge cell i stops once the local supply of either downstream cell has been exhausted. This assumption is reasonable for a diverge cell that has only one lane. However, a diverge cell with multiple lanes should permit some lanes to remain open while other lanes may be blocked with congestion. Neither of the diverge models described above can account for this partial blockage of a multi-lane diverge cell.

Networks with simple intersections can be modelled using a parallel implementation of the cell-transmission model. The density of a cell is independent of the density in other cells, and (partial) flow is independent of flows exiting other cells. Therefore, the exit flow of each cell can be calculated in parallel, as in Step 1 of Figure 2.14, then the density of each cell can be updated in parallel, as in Step 2 of Figure 2.14.

These simple intersection models can be generalised to accommodate merges of more than two upstream arcs and diverges of more than two downstream arcs. However, difficulties arise when using these merge and diverge models to model general intersections, in which a diverge cell commonly empties into a merge cell.

Lebacque (1996) proposed a simple model which combines the first merge model and the first diverge model—and their shortcomings—described above. The partial flow from any upstream (diverge) cell i to a downstream (merge) cell j is given by

$$\frac{y_{ij}}{\Delta t} = \min(\alpha_{ij}\tilde{q}_{D,i}, \beta_{ij}\tilde{q}_{S,j}).$$

Unfortunately, this simple model fails to maximise the usage of downstream local supply, nor can it capture the effects of blockages of upstream arc exits due to downstream congestion.

Ziliaskopoulos and Lee (1997) separated the diverge from the merge by using subcells in the last cell of each arc. The number of subcells matches the number of downstream arcs at the intersection. A diverge node upstream from the subcells separates the traffic by downstream arc. Each subcell carries a traffic stream for a single downstream arc. The outflow capacity of the roadway is shared among the subcells according to heuristic rules outlined in the Capacity Manual (*Special Report 209*, 1998). This reliance on heuristic rules is not a part of the cell-transmission model, as presented by Daganzo, but should be considered as an extension to the basic cell-transmission model.

2.4.3.2 Origins and destinations

The conservation law is permitted to be broken only at origins and destinations. Vehicles are created (released) at origins, and they are destroyed (absorbed) at destinations. Daganzo (1995a) proposed structures for origins and destinations.

An origin node is composed of two cells: a source cell followed by a gate cell. The source cell has infinite storage capacity, occupancy, and flow capacity. The gate cell has infinite storage and outflow capacity, but the inflow capacity matches the sum of the demand rates for all routes to all destinations that leave that origin over the next time step $[t_{j+1}, t_{j+2}]$

$$Q'_{gate,in} = \sum_{p \in \{o\} \times D} \sum_{r \in R_p} \int_{t_{i-1}}^{t_{i}} g_r(t)\,\mathrm{d}t\,.$$

The occupancy of the gate cell is generally equal to the inflow capacity, since it will usually discharge the entire occupancy during each time-step. However, if the occupancy is unable to fully discharge from the cell, then the remaining occupancy will be accumulated. This structure permits spillback at the origin without affecting the overall demand.

The destination is modelled as a sink cell with infinite storage capacity and inflow capacity, but zero outflow capacity.

2.4.3.3 Multi-commodity flow

Application of the cell-transmission model to the problem of dynamic network loading (DNL) requires some structural refinements to the cell occupancy. In DNL many temporal demand flows are assigned to particular destinations along specific routes. At any diverge along the route, the aggregate flow must be split among the alternative downstream arcs according to the specified routes. Since the arrival times at the diverge are unknown prior to the simulation, it is impossible to prespecify a temporal splitting rate that correctly partitions the aggregate flow to correspond to the specified routes. Therefore, the flow must carry a route label which can be used to identify the downstream arc at a diverge.

Some structure is necessary to track the multi-commodity flow. The multi-commodity flow, a term originally derived for flow-based static analytical models and network optimisation problems in operations research, refers to the origin-destination path flows which form the aggregate flow on the network. The term has since been used in a broader context of traffic modelling to mean any flows which cannot be mixed, such as flows of different vehicle types, flows with different priorities, or flows on different routes. These flows may share the same roadway, but the quantity of each flow must be known.

In the context of the traditional cell-transmission model multi-commodity flow refers to the various paths to each destination. Therefore, to implement multi-commodity "flow," the cell occupancies n^{t}_{idr} must be disaggregated by destination d and by route r. The cell occupancies must be further disaggregated by arrival time τ, i.e. a number of vehicles $n^{t}_{\mathrm{idr}\tau}$ on route r, which ends at destination d, entered cell i during the interval $[\tau, \tau + \Delta t]$, and remain in the cell at time t_j where $t_j \geq \tau$. In order to maintain FIFO discipline at the clock-tick level, vehicles that entered the cell earliest have departure priority over later arrivals. This disaggregation assumes that all vehicles which entered a cell at the same clock-tick τ did so at a constant flow rate over the entire interval $[\tau, \tau + \Delta t]$. Without the discretisation by arrival time τ, the new arrivals would be assumed to mix evenly over the length of the cell with the residual occupancy of the cell, which is a poorer approximation to FIFO discipline.

A similar occupancy disaggregation was implemented in NETCELL (Cayford, et al., 1997). The smallest unit of occupancy is called a packet. A packet with size $n_{\mathrm{id}\tau}(t)$ contains vehicles which remain on cell i at time t, entered the cell during time interval $[\tau, \tau + \Delta t]$, and are headed toward destination d. All packets on the cell with the same entrance time are grouped into a cohort. The cohort size is given by $n_{\mathrm{i}\tau}(t) = \sum_{\mathrm{d}} n_{\mathrm{id}\tau}(t)$. The cell occupancy $n_{\mathrm{i}}(t) = \sum_{\tau} n_{\mathrm{i}\tau}(t)$ is stored as a queue of cohorts in ascending order of entrance time to the cell. First-in-first-out discipline is strictly respected by taking cohorts from the head of the cohort queue to form the exit flow y_{i}^{t}. NETCELL permits multiple routes by specifying temporal route choice coefficients at each diverge to each destination.

Ziliaskopoulos and Lee (1997) implemented a different data structure for the modelling of vehicles discretely. A vehicle queue is represented with a one-dimensional array of pointers equal in size to the number of vehicles generated in the simulation. Each array element points to the vehicle ahead of it. Additional pairs of pointers mark the vehicles at the front and rear of each arc. The aggregate cell occupancies are used to calculate flow as usual, and in turn cell occupancies are updated. The pointer pairs are updated whenever flow exits an arc or origin and enters a downstream arc or destination.

2.5 Discussion of the Cell-transmission Model

The objective of this discussion is to qualitatively compare the cell-transmission model with other DNL models. Comparisons are made with other categories of DNL models, namely analytical models, mesoscopic models, and microscopic models, as well as other PDE approximations of the LWR model and higher order macroscopic models. Though practical concerns such as data requirements and calibration are included in this commentary, no quantitative comparisons are drawn. The comparison is based on the following criteria, some of which are borrowed from Khattak and Jovanis (1990):

- the soundness of the theoretical foundations of the model and the limitations of the assumptions which were used to apply the *theory* to create a practical *methodology*;
- the number and types of policy variables which are available to the model user for decision making, and the degree of *policy sensitivity* of the model;
- the *simplicity* of the model in terms of user comprehension and application;
- the types of *data requirements* and the degree of difficulty in obtaining observations; and
- the suggested *domain of application* based on resolution (or fidelity), accuracy, and computational efficiency.

2.5.1 Theory and methodology

Whereas the theory of analytical and macroscopic models is well documented in the academic literature, the description of microscopic and mesoscopic models is often vague and out-of-date. The accompanying manuals of commercially available models, such as INTEGRATION and AIMSUN2, are usually more current than the academic literature, but they are not always comprehensive in their description of the theory (perhaps due to their commercial nature). In addition, microscopic and mesoscopic models contain many heuristic rules which have little or no theoretical foundation.

The theories used to develop the analytical and macroscopic models often are based on restrictive conditions which imply simplifying assumptions to the DNL model.

Analytical models lack the structure and resolution to capture within-arc dynamics. The continuous representation of density and flow, and the absence of explicit lanes make modelling gap-acceptance impossible. Therefore, lane-changes, which are a form of gap-acceptance, do not reduce the capacity or cause any speed reduction to the traffic flow. Lane striping, which is the designation of permitted movements from each lane at the exit from an arc, is also ignored. So, gap-acceptance for opposed movements at intersections is ignored. The accuracy of analytical models is further compromised by the simplified flow-density relationship, as displayed in Figure 2.23. These crude modelling assumptions warrant the descriptor "low-fidelity" for analytical models.

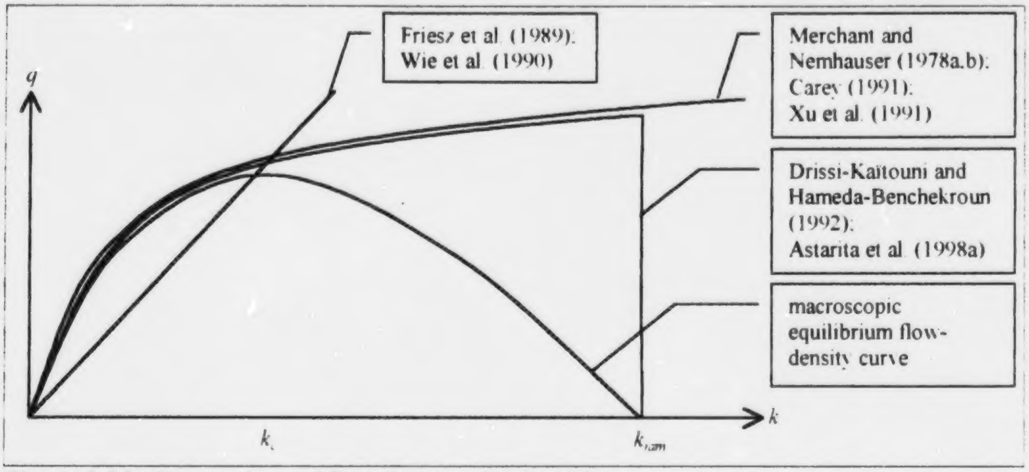


Figure 2.23 Comparison of flow-density relationships

The optimal control theory of Friesz et al. (1989) and Wie et al. (1990) necessitated the use of linear flow-density relationships. Clearly, this is a poor approximation. Firstly, densities which are higher than jam-density are permitted. Thus all queues are point queues, or vertical queues. Secondly, it is assumed that higher densities always imply greater exit flows. This assumption is obviously wrong for any density higher than the critical density. Therefore, these models cannot capture congested conditions, and the constant slope of the flow-density relationship in the uncongested regime makes any results from these models suspect.

Analytical models proposed by Merchant and Nemhauser (1978a, 1978b), Carey (1991), and Xu et al. (1999) employ a non-linear flow-density relationship which asymptotically approaches an exit flow capacity at infinite density. The presence of vertical queues implies that the effects of spillback are not captured. Furthermore, the non-decreasing nature of the flow-density relationship makes these models unable to properly model congestion.

The introduction of the jam-density parameter in models by Drissi-Kaïtouni and Hameda-Benchekroun (1992) and Astarita et al. (1998a) allows for the modelling of spillback. However, the non-decreasing nature of the flow-density function implies that queues are stored horizontally at jam density only. This dense queue will delay the onset of spillback. Shockwaves within the queue travel at infinite speed, which further delays or inhibits the start of spillback.

Macroscopic models also represent density with a continuous function, thereby ignoring the consequences of gap-acceptance. One improvement over analytical models is the use of an equilibrium flow-density relationship, which is permitted to decrease in the congested regime. This allows for a better description of the traffic in the congested regime.

In the LWR hydrodynamic model the flow state is restricted to the equilibrium flow-density relationship, which implies infinite vehicle acceleration and deceleration and no dissipation of shockwaves. Daganzo (1994, 1995a) suggests the use of a triangular or trapezoidal equilibrium flow-density relationship. This reduces the numerical error since all uncongested flow travels at the free-speed, jumping an entire cell at each time-step. However, fans are replaced by shockwaves which do not satisfy the entropy condition. Daganzo (1999) also proposes the use of a non-concave equilibrium flow-density curve, which may pose problems with convergence to the entropy solution of the LWR model.

In order to limit vehicle acceleration and deceleration, higher order macroscopic models permit the flow state to be outside of the equilibrium flow-density relationship by incorporating a reaction time—assumed to be zero in the LWR model— and an anticipation coefficient (Payne, 1971). However, the inclusion of these variables tends to lead to undesirable results, such as negative vehicle speeds and stalled queues when a signal changes from red to green (Lebacque and Lesort, 1999).

The LWR model ensures that small perturbations of the initial conditions of the system propagate forward in the uncongested regime, backward in the congested regime, and never travel with a velocity that is greater than the vehicle speed. Other PDE models may not share these characteristics. For example, second order PDEs may propagate disturbances in both directions simultaneously. Furthermore, the approximation employed by the cell-transmission model converges toward the solution of the LWR model for disturbances that travel either downstream or upstream, unlike other numerical methods, such as forward differencing schemes. (Daganzo, 1995b)

Shockwaves are dissipated due to discretisation in the cell-transmission model. This dissipation is greatest for densities equal to the critical density, but the dissipation does not occur at densities equal to zero or the jam-density. This dissipation is a purely numerical phenomenon and is not related to shockwave dissipation due to car-following and finite acceleration/deceleration. (Lebacque, 1996)

Another numerical error of the cell-transmission model—one which receives little attention in the literature—is the inaccuracy in the representation of the arc length due to the discretisation definition $\Delta x = \Delta t \cdot v_f$. An arc, which has a length that is not a multiple of the cell length, may have an error in length as large as $\pm \Delta x / 2$. For short arcs, the relative error in arc length, and consequently in arc travel time can be large. The error in arc length also affects the storage capacity of the arc, which affects the potential for spillback.

The cell-transmission model as defined by Daganzo (1995a) does not have the structure to model general intersections. Diverge cells cannot send flow directly to merge cells in order to preserve the property of a fully parallel implementation, which requires that the exit flow from a cell depend only on the density of that cell and cell(s) immediately downstream. Therefore, a diverge is required on each approach, upstream from the intersection, in order to separate the traffic streams on ordinary cells. A traffic stream may then flow from the ordinary cell into a merge cell on the downstream arc, and merge with traffic streams from other approaches. This representation of an intersection with merges and diverges is similar to a highway interchange.

Due to the absence of explicit lanes, partial blockages at arc exits with multiple lanes are modelled as either fully blocked or never blocked, according to cell diverge models described previously. Lane striping and gap-acceptance for opposed movements are ignored. Ziliaskopoulos and Lee (1997) described a modified cell-transmission model which used heuristics from the Highway Capacity Manual (Special Report 209, 1998) to model the effects of lane striping and gap acceptance for opposed movements.

2.5.2 Policy sensitivity

The DNL solution may be affected by many features of the network and the traffic. The usefulness of a DNL model in decision making depends on the number and type of policy variables available to the model user, and the sensitivity of the model output to these inputs. It is desirable to capture the effects of as many of these policy variables as possible in a comprehensive DNL model so that their effect on the arc flows and travel times can be evaluated. These policy variables can be separated into network features and fleet features.

The first group, network features, consists of attributes of the network $\widetilde{G} =< N, A >$, where N is the set of nodes, and A is the set of arcs. This group can be further subdivided into arc attributes and node attributes.

The arc attributes are reflected in the flow-density relationship of analytical and macroscopic models. Macroscopic models have the flexibility to use any temporally varying (or constant) concave flow-density relationship. So, they can be sensitive to the posted speed limit, arc flow capacity, arc storage capacity, the grade of the arc profile, and the presence of an incident on a portion of an arc. Analytical models have more restrictions which limit their ability to capture the effect on the arc travel times of the policy variables mentioned above. Many mesoscopic and microscopic models can also capture the effects of additional features, such as the length of weaving sections, the presence of lanes reserved for high-occupancy vehicles (HOVs), and various types of toll plazas.

Analytical models are capable of accounting for only the most basic node attributes, such as the exit capacity of the arc due to the intersection flow control. The cell-transmission model can explicitly model signal timing plans and ramp meters, though only as precisely as the time discretisation will permit. Actuated and adaptive signal timing can also be reproduced by the cell-transmission model, by linking the flow capacity of the last cell to the occupancy of upstream cells, though only as precisely as the space discretisation will permit. Lane-striping, roundabouts and permissive phases for left turns can only be crudely modelled by the cell-transmission model. These features and others, such as roundabouts and stop signs, are modelled far better by mesoscopic and microscopic models, due to their discrete vehicle, discrete link simulations.

The second group, fleet features, relates to the ability of the DNL model to account for a heterogeneous fleet of vehicles and drivers. Mesoscopic and microscopic models have the flexibility to model the explicit presence of trucks and pedestrians, the varying acceleration and braking capabilities of all vehicles, and variability of driver behaviour in terms of aggressiveness. Analytical and macroscopic models are restricted to using heuristic rules and passenger car equivalencies to capture any of the above fleet features.

2.5.3 Simplicity

Analytical models make use of the most abstract representation of traffic and the network. The user often requires knowledge of specialised fields of mathematics to understand the policy sensitivity.

The theory used to develop macroscopic models is arguably more complex than analytical models, but the implementation hides these mathematical complexities from the user. Instead, graphical concepts can be used to explain the functionality of the cell-transmission model. The most difficult graphical concept is characteristics, since their trajectory follows kinematic waves, which are an abstract notion. The inclusion of a flow-density diagram next to any space-time diagram greatly aids the reader to gain an intuitive understanding of the traffic flow dynamics.

Microscopic and mesoscopic simulation models use more realistic representations of the network and traffic. Hence, the model user has a more intuitive grasp of the methodology of the DNL model. The number of parameters gives the model user a great deal of flexibility when calibrating the model. Though this flexibility is beneficial to a user who understands how and why the parameters affect the model output, less informed users may be disoriented by the dizzying array of options.

2.5.4 Data requirements

Obviously, analytical models require the least amount of data, since they use the most abstract representation of networks and have the least policy sensitivity. Often an arc-node incidence matrix, the arc lengths, the number of lanes, and the necessary parameters which define the exit link function or travel time function suffice to obtain a solution to the DNL problem.

Macroscopic models require additional arc parameters to define the flow-density relationship. Detailed signal timing plans are also useful.

Microscopic models are data intensive. They may require additional arc attributes, such as an estimate of the variance of the arc parameters. A virtually limitless array of data may be required. The following list names a few of the additional data requirements of the microscopic traffic simulation program called INTEGRATION: node macro-zone cluster numbers, vehicle speed variability, lane striping, and critical gap sizes. Default values for many parameters are often suggested in the user manual.

2.5.5 Domain of application

The suggested domain of application is based on the resolution (fidelity) of the model and on the computational efficiency.

Analytical models are relatively crude in terms of accurately modelling traffic dynamics and policy sensitivity. In fact, congestion is either modelled poorly or not at all. The models can be solved by analytical methods or simulations with minimal data requirements. They are most appropriate for planning purposes, which generally require crude estimates of travel times on networks larger than 1000 arcs.

Macroscopic models can adequately capture traffic dynamics in uncongested and congested conditions. Therefore, they are useful for planning purposes when greater accuracy is necessary or when congested conditions prevail. They are particularly well suited for intelligent transportation system (ITS) design, implementation testing, and evaluation of benefits. Most ITS initiatives require real-time—or even predictive—knowledge of travel times and traffic volumes. ITS initiatives often involve exploitation of alternative routes in an integrated network of freeways and surface streets. Therefore, the solution to the dynamic network problem must be fast, sensitive to subtle network features and traffic phenomena, and applicable to large networks. Though macroscopic models have less policy sensitivity compared with mesoscopic and microscopic models, they can provide (predictive) travel time estimates more quickly and they can model larger networks.

The cell-transmission model has two additional features. Firstly, it is well suited for parallel implementation which may produce faster simulation speeds. In addition, the cell-transmission model offers the flexibility to capture within-arc shockwaves and spillback at intersections with a precision determined by the degree of discretisation that may range from coarse to very fine. Of course, increased precision comes at the cost of increased computational demands. Thus, the cell-transmission model is a good tool for deciding what level of precision is necessary for particular networks and temporal demands.

Though microscopic models and mesoscopic models are capable of simulating urban areas on a personal computer, the *real-time* simulation of large, congested networks requires far more computer processing power. Their greater policy sensitivity and modelling realism come at the price of computational expense. These models excel at accurately simulating small networks of only a few intersections, though they are also appropriate for site impact studies and corridor analysis on networks smaller than 1000 arcs.

2.5.6 Implications of theory on practical applications

The cell-transmission model was initially conceived as a finite difference approximation to the first order partial differential equations which describe the LWR hydrodynamic model. During the writing of this chapter, it became clear that the links between the models are complex. A comprehensive understanding of the relationship between the models requires in-depth knowledge from the fields of fluid dynamics, thermodynamics, numerical methods, traffic flow theory, and traffic engineering. Consequently, specialists in each field developed and documented only parts of the relationship between the models. No comprehensive explanation which united these contributions was found in the literature.

Nevertheless, researchers have adopted the cell-transmission model to solve the DNL problem (Sawaya et al., 1999; Li et al., 1999, Lo, 1999). Adjustments to the cell-transmission model were made with little concern over the basic assumptions which are necessary for demonstrating convergence to the physically relevant solution of the LWR model. Some of these adjustments were intended to improve policy sensitivity to a heterogeneous fleet of traffic on a realistic network with general intersections. Others were introduced to reduce the computational demands of the solution algorithm.

One example of this engineering approach is the use of the cell-transmission model in conjunction with a piecewise-linear flow-density relationship. Daganzo (1994) suggested using the piecewise-linear equilibrium flow-density relationship in the cell-transmission model without first discussing the impact of such a flow-density relationship on the LWR model. After studying the LWR model from several perspectives, it was discovered that the LWR model with a piecewise linear equilibrium flow-density relationship has a unique solution, which does not respect the entropy condition or driver's ride impulse. This unique solution contains additional shockwaves which do not appear in the entropy solution for smooth equilibrium flow-density curves. The solution of the cell-transmission model converges to this unique solution of the LWR model with finer time-space discretisations.

Another example, in which the cell-transmission model was altered without justification, can be found in Ziliaskopoulos and Lee (1997). The authors described an implementation of the cell-transmission model with discrete vehicles, and a network that is discretised finely for surface streets and coarsely for freeway arcs. Boundary conditions must be specified at the freeway ramps to buffer the flow between the long freeway cell and the short arterial cell. An examination of the impact of these boundary conditions on the approximation of the LWR solution is necessary to justify the variable discretisation of arcs. The movement of discrete vehicles to satisfy the real-valued flows prescribed by the cell-transmission model also poses an interesting implementation problem.

Many such variants of the cell-transmission model are worth consideration in order to efficiently and accurately solve the DNL problem for realistic networks with a wide variety of policy variables. The following alterations to the cell-transmission model are suggested for study:

- relaxation of the discretisation constraint $\Delta x = \Delta t \cdot v_f$ for one cell per arc, so that the arc length can be represented accurately;

- the introduction of a buffer between arcs with different time discretisations, as described above by Ziliaskopoulos and Lee (1997);
- an improved node model for general intersections—not simple merge and diverge intersections—to capture partial blockages at the arc exit for arcs with multiple lanes and explicit lane striping, which addresses the implications of first-in-first-out discipline;
- the modelling of cells which function as both a merge cell and a diverge cell, so that short arcs may be represented using a single cell, instead of halving the time discretisation and effectively doubling the computational demands of the solution algorithm;
- relaxation of the deterministic flow-density relationship to represent the approximate nature of the estimation of the equilibrium flow-density relationship from empirical observations; and
- the implementation of schemes which integrate heuristics that represent the impacts of features of traffic dynamics that are too fine to be explicitly modelled by the cell-transmission model, such as the discrete vehicles engaged in gap-acceptance during lane-changing manoeuvres and at intersections with permissive movements.

These alterations must be examined for theoretical soundness. The objective is to extend the cell-transmission model to account for more policy variables with increased policy sensitivity without abandoning or contradicting the theoretical framework which justifies the convergence to the LWR model. These alterations will also be considered from the perspective of the implementation of the algorithm and its computational efficiency, along with the quantitative impact on the temporal travel time estimates.

The study of these alterations will be performed within the context of an efficient implementation of the cell-transmission model in an object-oriented programming language is detailed in chapter

3. In chapter 4, several small demonstration networks are modelled to validate the program code. Results are compared to the output from a microscopic traffic simulation model, which serves as a benchmark to evaluate the accuracy of the cell-transmission model. The alterations and extensions to the cell-transmission model are considered in chapters 5 and 6. Chapter 5 addresses flow propagation, while issues concerning node models are examined in chapter 6. The performance of the enhanced cell-transmission model in solving the DNL problem for traffic networks is analysed in Chapter 7.

Chapter 3

Implementation of the Cell-transmission Model

3.1 Introduction

This chapter describes the implementation of the cell-transmission model in a simulation program, called CellNetLoad. The structure of the program was generalised to permit testing of the basic cell-transmission model, as well as versions which include enhancements. A single program code was written for the basic model and the enhanced models to ensure that the differences in the simulation results were attributed to the model enhancements and not implementation details which differed in each program. The enhancements are activated using flags that are provided as input to the program.

Computational performance was sacrificed for the sake of flexibility, because some of the model enhancements required additional data structures and methods than the basic cell-transmission model. Therefore, relative comparisons of the computational performance must consider that the basic model is carrying some overhead in terms of inefficient data structures and methods used only when the model enhancements are activated. In addition, only crude estimates could be derived for the absolute computational performance of the cell-transmission model as a solution to the dynamic network loading problem.

In the remainder of this chapter, the framework of the simulation program and the implementation of the logic of the basic cell-transmission model are documented. The enhancements are discussed in later chapters. Flow charts are used to convey program flow. The most complex program logic is described in the form of an algorithm.

3.2. Development context

Java was selected as the implementation language for several reasons. The complexity of modelling a traffic network with the cell-transmission model warranted the use of an object-oriented language, which provides the interfaces to write modular code. Java is a well-documented introductory language for object-oriented programming. It is also appropriate for prototyping, which was necessary during development, testing, and refinement of the model enhancements. Both a UNIX network and Windows personal computers were available for this project, so the platform independence of Java was a welcome feature. In addition, NetBeans DeveloperX2 2.1, a full-featured, cross-platform integrated development environment for the Java 2 platform was available free of charge for download, and proved to be a remarkably stable and effective program.

CellNetLoad was written as a single thread. Because the cell-transmission model was conceived as a model which could benefit from a parallel implementation, the option to use multiple threads and multiple processors was an obvious consideration. Both the Java language and NetBeans DeveloperX2 support multiple threads. So, future research with parallel implementations is possible without switching to a different development environment. One drawback of the Java language is that the cross-platform flexibility comes at the price of increased program execution times. As explained above, this implementation is used only to draw conclusions about the relative performance of the basic cell-transmission model and enhancements. Therefore, the absolute program execution speed was not of crucial importance for this research.

3.3 Network representation

Before explaining the representation of the network, the relevant data structures should be briefly described. The arcs, nodes, origins, destinations, and signals which constitute the network are stored as objects in tables. The table indices serve as the identification numbers for each object. The identification numbers are not required to be sequential, but they are limited by an upper bound. The tables and the upper bounds on their size are stored as static variables in the Simulation class.

The Node class contains x- and y- co-ordinates, an origin identification number if it is an origin, a destination identification number if it is a destination, the arc identification numbers in the forward star and backward star, and several arrays which store data used to distribute local demand and local supply at the intersection. These arrays are discussed later.

The Arc class contains the identification numbers of the upstream and downstream nodes, the variables which describe the arc itself, a list of cells which contains the flow on the arc, a table of path information, and an object used for the collection of statistics for output. The arc is described by the length and the number of lanes; free speed, speed-at-capacity, flow capacity, jam density, and wave speed; and the identification number of the signal which controls the outflow and those phases of the signal timing plan which are green. The path table contains the identification number of the next arc in the path given the identification numbers of the destination node and the path.

The Signal class contains signal timing data, including the cycle length, the offset, the number of phases, and the duration of each phase. The signal is not tied to any specific node or arc. Instead, each arc may use the signal timing of any one signal, and discharge on any of the phases of that signal.

The Cell class has descriptor variables and flow variables. The cell is described by a maximum occupancy, a flow capacity, a density factor, and a flow factor. The factors are used to reduce the maximum occupancy and the flow capacity during an incident. Flow is stored in the cohort queue while it occupies the cell. Cohorts ready for departure from the cell are temporarily stored in a second queue.

The LastCell class is a subclass of Cell. Each arc contains a list of cells. The last cell in the list of cells is a LastCell. In addition to the member variables of the superclass Cell, the LastCell class has a signal factor, which is applied to the flow capacity to determine the cell outflow capacity. The LastCell class also contains a list of secondary queues to store the cohorts ready for departure to each arc in the forward star.

The Cohort class contains a time stamp, a list of packets, and the size. The time stamp stores the time of the simulation clock when the cohort entered the arc. The size of the cohort is the total number of vehicles in the list of packets.

The Packet class contains the identification numbers of a single destination node and a single path, as well as the size. The size of the packet is the number of vehicles following the path to the destination.

The Origin class contains several cohorts and lists to load demand, a table of path information, and an object used for the collection of statistics for output. The variables used to load demand are described later. The path table contains the identification number of the first arc in the path given the identification numbers of the destination node and the path.

There is no Destination class. Packets and cohorts which exit the network are not stored. They are recovered during garbage collection.

The Statistics class stores the number of entering vehicles, the number of exiting vehicles to each arc in the forward star, the accumulated occupancy, and the total travel time. The manipulation of these statistics to generate output data is explained later.

Though some of the data structures contain redundant information, the lack of efficiency was tolerated because it made the development of model enhancements simpler. If computational efficiency were to be evaluated absolutely, then these redundancies would first be removed.

In summary, Packets are contained by Cohorts, which are contained by Cells (and LastCells), which are contained by Arcs, which are stored in a table, which is a static variable in the Simulation class. Nodes, Origins, and Signals are stored in separate tables, each of which is a static variable in the Simulation class.

Arcs are modelled as a single lane with arc flow capacity and arc jam density equal to the per-lane values multiplied by the number of lanes. The arc is discretised along its length into cells. The number of cells used to represent an arc depends on the time-step parameter, which is provided as input and stored as a static variable in the Simulation class. The length of a cell is equal to the distance travelled on the arc at free speed during one time-step. Arcs are represented by an integer number of cells. Consequently, the time-step must be smaller than the shortest arc travel time at free speed, since an arc can be no shorter than one cell. Furthermore, the arc length is not modelled accurately: it may be up to one cell length too short. Arc storage and arc travel times are underestimated due to this modelling error.

The output of the program represents the approximation of the cell-transmission model to the LWR solution. The approximation improves with a smaller time-step. Theoretically, the approximation is exact for an infinitely small time-step. In practice, the program execution time (at least) doubles as the time-step is halved.

The basic cell-transmission model does not model intersections with multiple arcs in the backward star *and* the forward star. The generalised intersection model of CellNetLoad permits multiple arcs in the forward star and backward star. A stream may diverge from other streams as it exits the last cell of an arc in the backward star, then immediately merge with streams from other arcs as it enters the first cell of an arc in the forward star. In order to remain consistent with the spirit of the cell-transmission model, intersections are modelled as points, i.e. they have no physical length or width. The interactions of the streams at the intersection are governed by an iterative algorithm implemented and stored in the Node class.

The cell-transmission model represents the traffic stream as a fluid. Thus, a quantity of vehicles may be specified as a real-valued number. This creates a problem for implementation since computers express real numbers with limited precision. In order to avoid losing or gaining vehicles due to cumulative rounding errors, an epsilonic value is specified in the Simulation class. Whenever a quantity of vehicles is split into two fractions, each fraction is rounded to the nearest multiple of the vehicle epsilon. This solution succeeded in conserving flow, provided that the vehicle epsilon was sufficiently large. A value of 10^{-5} vehicles was found to be appropriate.

3.4 Input-output structure

The input and output interfaces were constructed to expedite verification and validation of the code by the developer. Any commercial use of this software would necessitate improvements to the input-output interfaces.

The input is provided to CellNetLoad via a set of text files and several command line arguments. This input structure was preferred over a graphical interface because it facilitated running many series of simulations for networks with varied inputs. This process was automated using script files.

The format of the text files was borrowed from the INTEGRATION user's manual. INTEGRATION is a microscopic traffic simulation program. A master file contains the duration of the simulation, and the file names of a node file, an arc file, a signal timing file, a demand file, an incident file, and a path file. The format of each of these files is provided in Appendix A. Only the relevant data was extracted from the text files by CellNetLoad. All extraneous data is ignored.

In addition to serving as a model for the input interface, INTEGRATION was used as a benchmark to validate CellNetLoad. The simulation results are compared in chapters 4 through 7. INTEGRATION proved to be a useful tool for validating this macroscopic simulation program because the car-following logic is based, in part, on macroscopic speed-flow-density parameters. Therefore, no calibration was necessary between the microscopic car-following model and the macroscopic flow propagation models. The movement of vehicles in INTEGRATION differs from the macroscopic models mainly in that vehicles are modelled discretely, with finite acceleration and deceleration. Both mandatory and discretionary lane-changing, as well as gap acceptance of opposed movements at intersections, are modelled explicitly by INTEGRATION.

The car-following logic used in INTEGRATION is based on a non-linear equilibrium speed-headway relationship defined by four parameters: the saturation flow rate per lane $q_{\max}$, the free speed v_f, the speed-at-capacity v_c, and the jam density per lane k_{jam}. These same parameters can be used to construct an equivalent non-linear equilibrium flow-density curve, shown in Figure 3.1. The cell-transmission model uses a piecewise linear equilibrium flow-density relationship.

Therefore, a trapezoidal relationship was fit to the non-linear flow-density curve using the given parameters. The slope of the line in the uncongested regime is equal to the free speed. The flow capacity is equal to the saturation flow rate. The slope of the line in the congested regime, called the backward wave speed w is specified as

$$w = \frac{q_{\max}}{k_{jam} + \frac{q_{\max}}{v_f} - 2\frac{q_{\max}}{v_c}}. \tag{3.1}$$

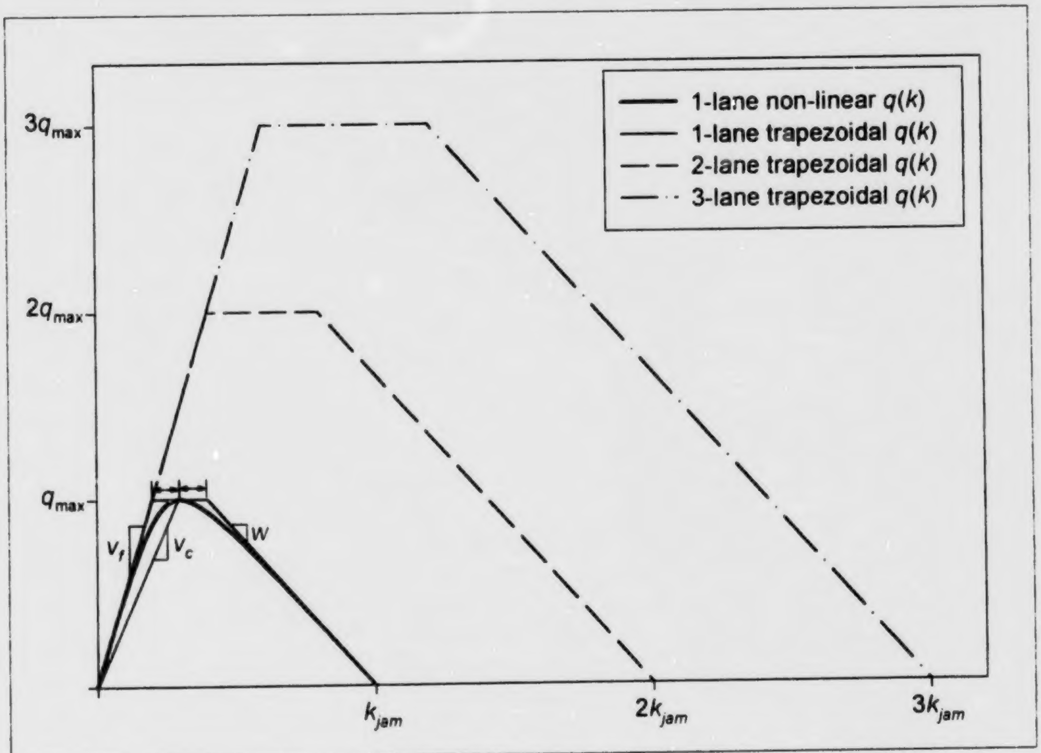


Figure 3.1 Piecewise linear equilibrium flow-density relationship used in CellNetLoad

This specification has the interesting properties that the density-at-capacity k_c lies at the midpoint of the horizontal line, and that the sloped line in the congested regime is tangent to the curve at the jam density, as shown in Figure 3.1. Both axes can be scaled up to account for multiple lanes without affecting the sloped lines which represent speeds.

One deviation from the INTEGRATION format was necessary for the incident file. Instead of specifying lane blockages with a bitstring, a fractional value is used to indicate the portion of the arc width which is blocked during the incident.

Command line arguments are used to specify the time-step, the duration of the aggregation interval for output data, the size of a vehicle unit, an eight-character string that activates particular enhanced features of CellNetLoad, and the location and name of the master file. The usage of the vehicle unit size is documented in chapter 5. The default value for the vehicle unit size is 0.001. Each character in the string is a flag. The function of each flag is briefly described in Table 3.1. The values of the flags for the basic cell-transmission model are in bold. More detailed descriptions of the implementation of the enhanced features are included in chapters 5 and 6.

Like the input data, output data also takes the form of text files. These text files were easily imported to spreadsheets using scripts. One text file is used for each of the following output statistics: arc inflow, arc outflow, arc partial flow, arc density, and arc travel time. Each output interval adds one row to each file. The beginning and end of the output interval are listed in the first and second columns. The data for each arc is stored in subsequent columns. The arc inflow and arc outflow are simply counts of the number of vehicles that enter and exit each arc during the output interval, divided by the duration of the output interval. The partial flow is the outflow separated by each arc in the forward star. The file containing partial flows has more columns than the other output files. The arc density is the average occupancy of the arc during the output interval, divided by the modelled arc length and the number of lanes. (The true arc length may be longer.) The arc travel time is the average arc travel time experienced by all vehicles which *enter* the arc during the output interval. The determination of the arc travel time by entrance time requires post-processing of the statistics. Post-processing is explained later.

Table 3.1 Activation of enhanced features of CellNetLoad using the command line flags

Position	Flag name	Description
	Flag value	Function
1	`exactArcStorage`	determines the representation of the arc storage by cells
	0	each cell contains an equal storage capacity
	1	the last cell accounts for the error in storage
2	`flowDensityModel`	determines the equilibrium flow-density relationship
	0	trapezoidal relationship
	1	non-linear relationship used by INTEGRATION
3	`signalModel`	determines the representation of signals
	0	static signals
	1	explicit phasing
4	`nodeModelType`	depricated
5	`nodeModelSplits`	determines the distribution of the local demand and supply
	0	proportionally to partial demand
	1	proportionally to cell discharge capacity
6	`fifoDiscipline`	determines the enforcement of first-in-first-out discipline
	0	FIFO is relaxed at the cell level
	1	FIFO is relaxed for all cohorts in the local demand
	2	FIFO is relaxed for all vehicles in the local demand
	3	FIFO is relaxed within each cohort only
7	`vehicleUnitRounding`	determines the treatment of partial vehicle units
	0	always round down
	1	use probabilistic rounding

3.5 CellNetLoad simulation program

CellNetLoad is an event-based simulation. The simulation could have been implemented using a loop because the movement of vehicles takes place at equally spaced time-steps. However, an event-based simulation was preferred because it permits flexibility in modelling events which may not fall exactly on the time-step.

The program execution is divided into three phases: initialisation, simulation, and post-processing.

3.5.1 Initialisation

During initialisation the input files are read. From the data in the input files, the binary data structures are built. The simulation terminate event is scheduled as the simulation duration is read from the master file. As each signal timing plan is read from the signal timing file, a signal timing event is scheduled for each signal in the event queue. The start and end times of each origin-destination demand in the demand file are scheduled in the event queue as demand events during the reading of the demand file. The start and end of each incident are also scheduled in the event queue as incident events while the incident file is read. One route event is scheduled for each route plan read from the path file.

3.5.2 Simulation

The simulation phase first schedules the initial clock tick event and output statistics event at time zero. Each event is processed according to its scheduled time. The following is a description of the process method for each event.

Each type of event is a subclass of the Event class. The Event class contains an execution time, an object which stores data, and a method which is used to process the data. Each subclass overrides the process method. Polymorphism allows the Manager class to process events without distinguishing one event type from another.

3.5.2.1 Route event

A route event contains a route plan, which specifies all the available paths to each destination and the proportion of demand on each path. During execution of a route event, the path tables in each arc and origin are updated, and the path splits stored in the Simulation class are updated. Vehicles en-route divert to the new paths.

It is recognised that the specification of a route plan is restrictive. The same path splits apply to all destinations. If different path splits are required for different destinations, then many paths must be repeated. For example, suppose there are up to two paths to each destination. If only one path exists to a particular destination, then it is stored twice. Now, suppose that demand to destination 1 is split equally on two different paths. Demand to destination 2 is split such that one quarter takes one path, and the rest follows a second path. This situation would require specification of 3 paths, one quarter on path 1, one quarter on path 2, and half on path 3. Paths 1 and 2 to destination 1 would be the same, while path 3 would be different. Paths 2 and 3 to destination 2 would be the same, and path 1 would be different. Though the route plans were sufficient for small networks and very basic tests, these rather cumbersome data structures must be revised before CellNetLoad could be used to solve dynamic equilibrium assignment problems.

3.5.2.2 Demand event

The demand event contains origin and destination identification numbers, as well as a flow rate expressed in vehicles per hour. The process method is depicted in Figure 3.2. The flow rate is multiplied by the time step to arrive at a quantity of vehicles which should depart the specified origin during each time step. This quantity is then divided among packets. The packets are each labelled with a different path identification number and the specified destination identification number. These packets are filled according to the path split coefficients, which are stored in an array in the Simulation class. The packets are then added to the demand cohort in the specified origin. Note that demand flow may be added and removed during the course of a simulation, so negative flow values are used to remove flow from the demand cohort.

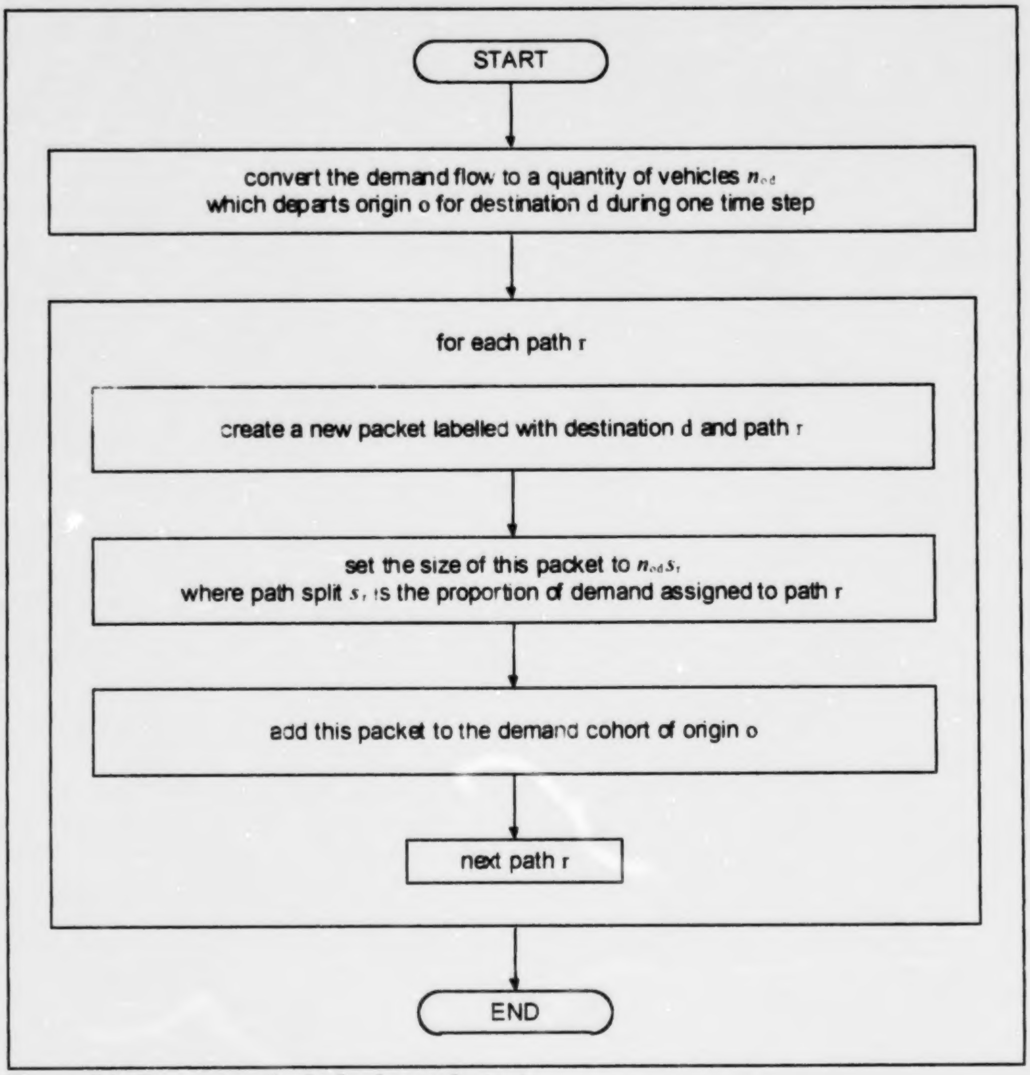

START
convert the demand flow to a quantity of vehicles n_{od}
which departs origin o for destination d during one time step
for each path r
create a new packet labelled with destination d and path r
set the size of this packet to $n_{od}s_r$
where path split s_r is the proportion of demand assigned to path r
add this packet to the demand cohort of origin o
next path r
END

Figure 3.2 Process method of a demand event

3.5.2.3 Signal event

The signal timing event contains a table of signals. The process method of the signal timing event is depicted in Figure 3.3. When a signal event is processed, the signal table stored in the Simulation class is updated. Then the signal factor is updated for the last cell in each arc.

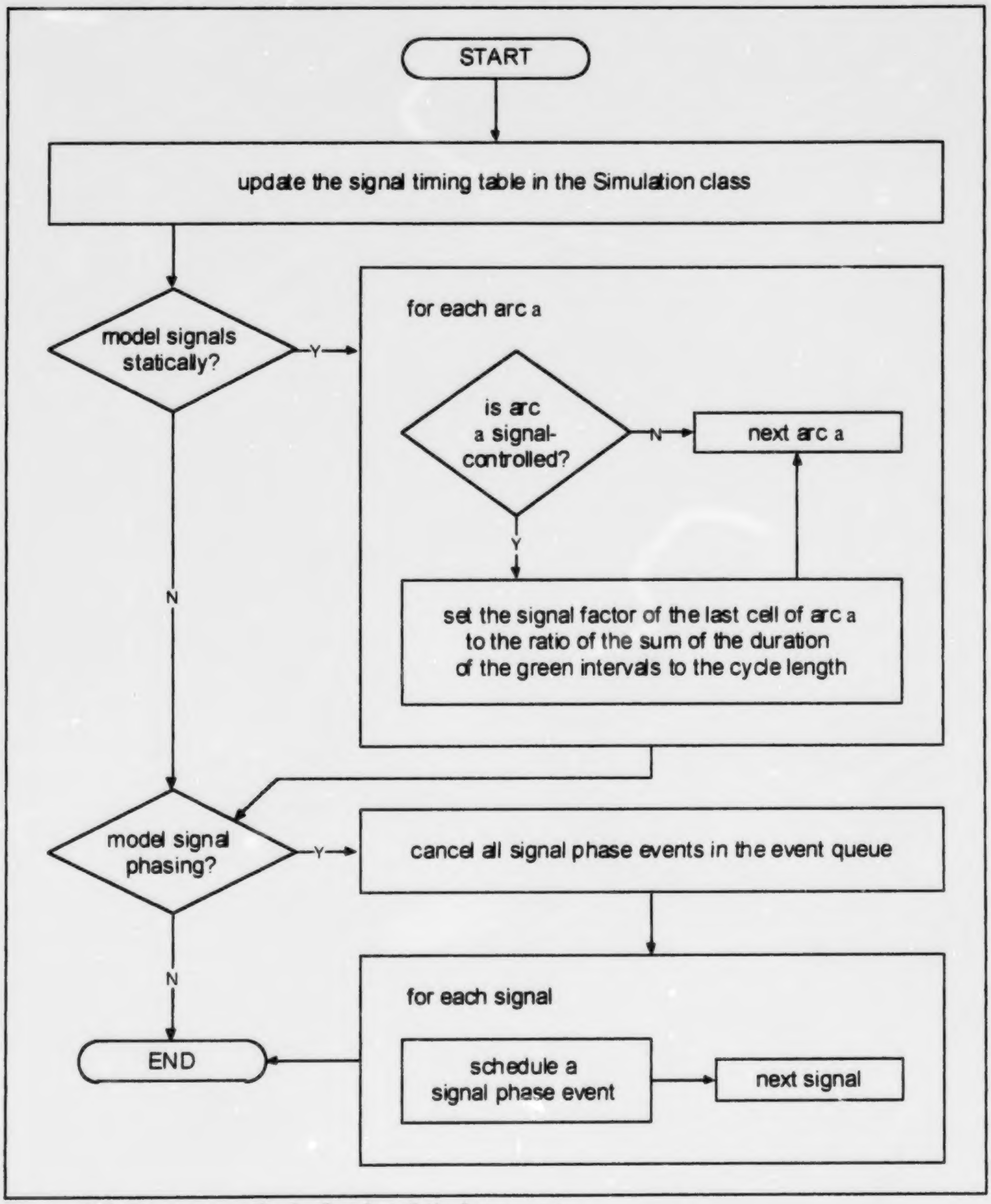

START
update the signal timing table in the Simulation class
model signals statically?
Y
N
for each arc a
is arc a a signal-controlled?
N
next arc a
Y
set the signal factor of the last cell of arc a to the ratio of the sum of the duration of the green intervals to the cycle length
model signal phasing?
Y
cancel all signal phase events in the event queue
N
for each signal
schedule a signal phase event
next signal
END

Figure 3.3 Process method of a signal timing event

If signals are modelled statically (`signalModel = 0`) then the signal factor is the ratio of the sum of the green intervals to the cycle length. If signals are modelled with explicit phasing (`signalModel = 1`) then the signal factor must alternate from 1.0 during the green interval of any discharge phase, to 0.0 during all other phases. This is accomplished by immediately scheduling a signal phase event for each signal.

A signal phase contains only the identification number of a signal. The process method of the signal phase event is depicted in a flow chart in Figure 3.4. When a signal phase event is processed, the signal factor in the last cell of all arcs controlled by that signal is updated. The number of the phase(s) for which an arc receives a green indication is stored in each arc. The duration of each interval of each phase and the offset are stored in the signal table in the Simulation class. If the current phase of the signal is a discharge phase for the arc, then the signal factor is set to 1.0, provided that the current simulation time falls within the green interval of that phase. If the current simulation time is within the yellow interval of the green phase or if the arc does not receive a green indication during the current phase, then the signal factor is set to 0.0. A new signal phase event is always scheduled for the end of the current interval of this signal.

3.5.2.4 Incident event

An incident event contains the arc identification number on which the incident occurs, and the start and end locations of the incident, measured from the upstream end of the arc, as well as the severity of the incident. The severity has a value between 0.0 (total arc closure) and 1.0 (no impact). The process method of an incident event traverses the cell list of the arc. If any part of the cell is between the start and end location of the incident, then the flow factor is reduced by the incident severity. The density factor of the cell is reduced by the product of the severity and the proportion of the cell which lies between the start and end locations of the incident. The effects of multiple incidents, which affect the same cell, are cumulative. The incident event which marks the end of an incident is identical to the incident event which marks the beginning of an incident, except that the severity is larger than 1.0, so that the flow factor and density factor will be increased.

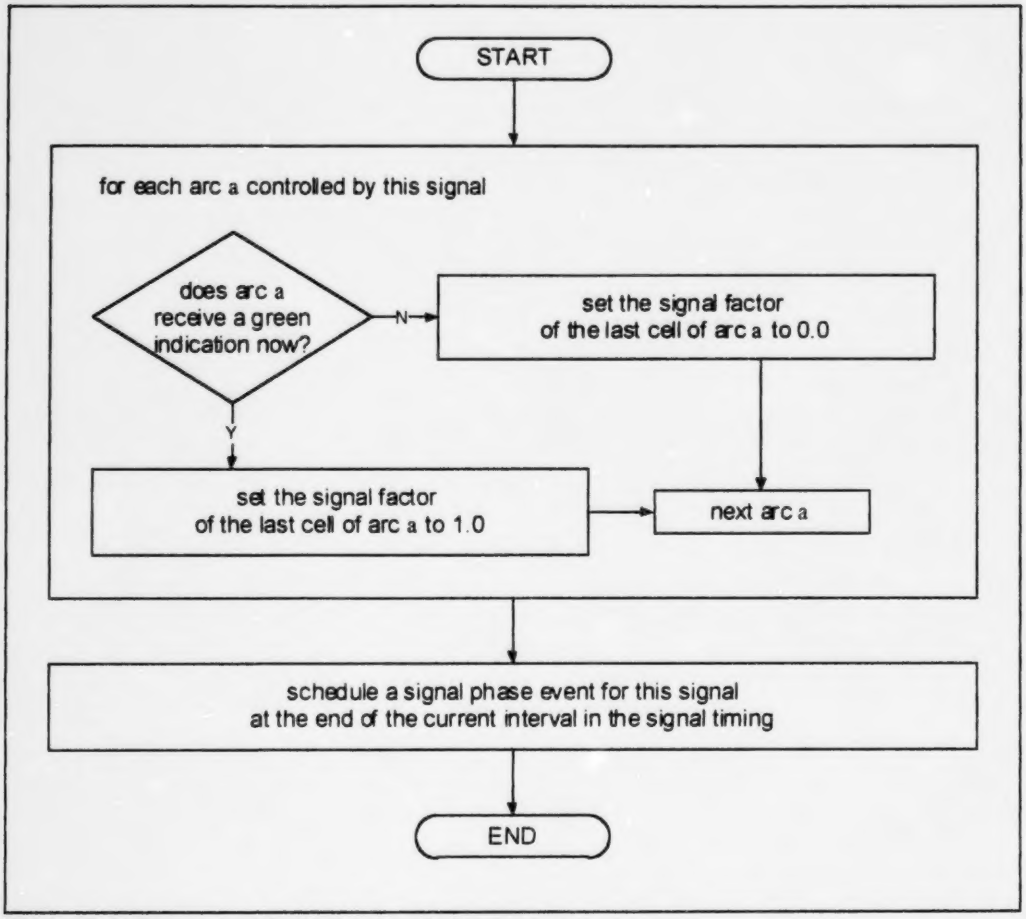

START
for each arc a controlled by this signal
does arc a receive a green indication now?
N
set the signal factor of the last cell of arc a to 0.0
Y
set the signal factor of the last cell of arc a to 1.0
next arc a
schedule a signal phase event for this signal at the end of the current interval in the signal timing
END

Figure 3.4 Process method of a signal phase event

3.5.2.5 Clock tick event

The clock tick event is the most complex type of event in CellNetLoad. The process method of the clock tick event is responsible for all flow propagation during the current time step, which begins at the scheduled time of this clock tick event, and which ends when the next clock tick event is processed. Flow is propagated from origins to destinations through the cells of arcs and across nodes. All necessary statistics are accumulated. The process method of the clock tick event is depicted in several flow charts.

The first task of the clock tick is to scan each node, preparing the correct number of vehicles to exit the last cell of each origin and arc, as shown in the flow chart of Figure 3.5. The treatment of the node depends on its status as an origin, a destination, both, or neither. An origin must have a single arc in the forward star. A destination node must have a single arc in the backward star. One node may be both an origin and a destination. A node that is neither an on origin or a destination, may have up to seven arcs in each of the backward and forward stars.

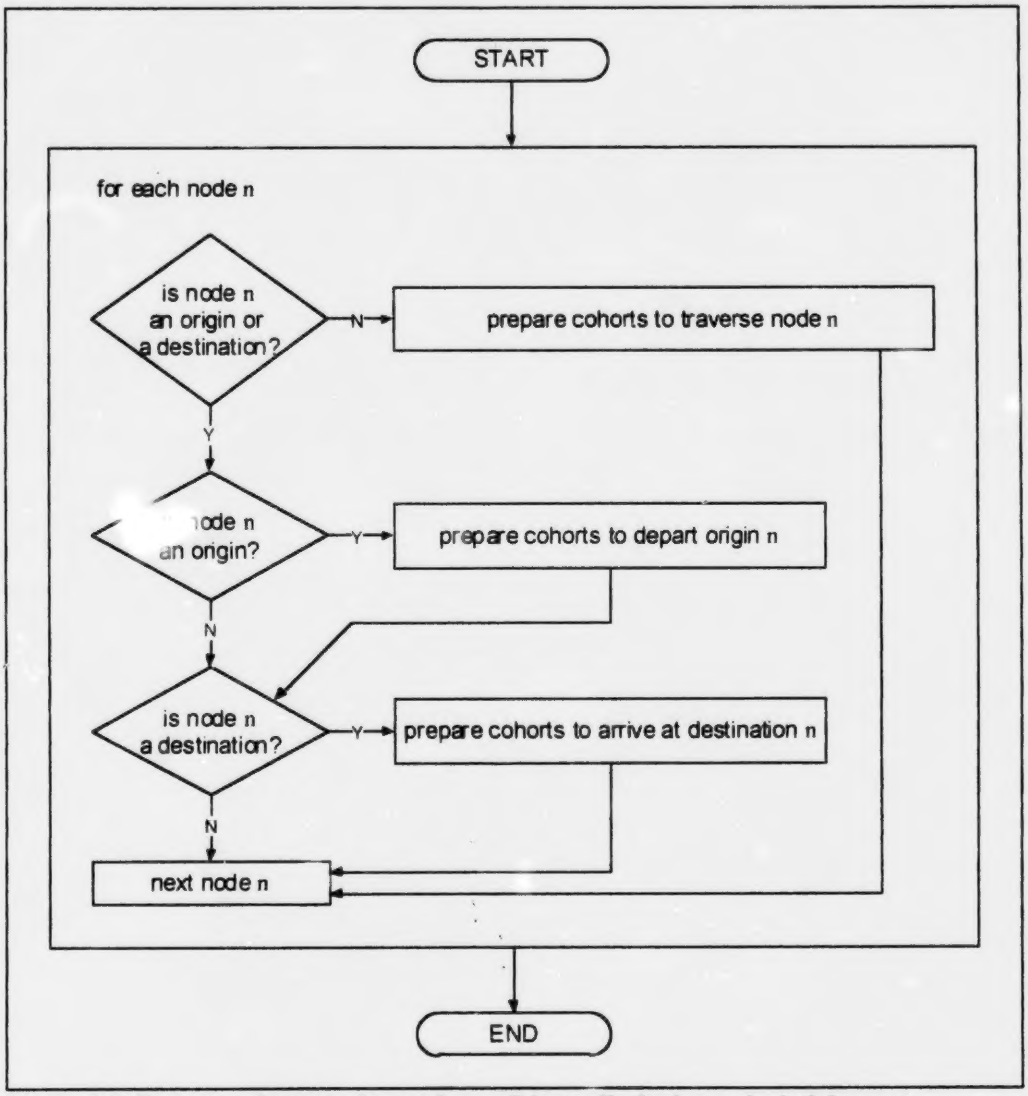

START
for each node n
is node n an origin or a destination?
N
prepare cohorts to traverse node n
Y
is node n an origin?
Y
prepare cohorts to depart origin n
N
is node n a destination?
Y
prepare cohorts to arrive at destination n
N
next node n
END

Figure 3.5 Prepare cohorts to depart from all last cells during a clock tick event

If the node is an origin, then flow must be generated and prepared for departure from the last cell in the origin, as shown in Figure 3.6. The data structures of the origin allow only integer multiples of the specified vehicle unit to enter the network. The basic cell-transmission model requires a small vehicle unit size to approximate continuous flow. An enhanced feature of CellNetLoad is the ability to discretise flow into units of any (fractional or integer) number of vehicles. The data structures of the origin also permit spillback to occur at the origin.

The source of flow is the demand cohort. Each packet of the demand cohort is cloned and added to the latent demand cohort, which serves to accumulate fractional vehicle units of demand. An integer number of vehicle units is removed from each packet of the latent demand cohort and added to a new cohort, which is then stamped with the current simulation time. This new cohort is the demand generated during this time step. The new cohort is added to the back of the cohort queue of the cell in the origin. This cell is an instance of the subclass LastCell, so it may be referred to as the last cell of the origin.

The number of vehicles which exit the last cell of the origin during this time step is the minimum of the local demand and the local supply. The local demand is calculated as the occupancy of the last cell of the origin, since it has infinite flow capacity. The local supply is

$$Q_{S,\mathrm{j}}^{j} = \begin{cases} q_{\max,\mathrm{j}}\Delta t & \text{if } K_{\mathrm{j}}^{j} \le k_{c,\mathrm{j}} \\ \tilde{q}_{\mathrm{j}}(K_{\mathrm{j}}^{j})\Delta t & \text{if } K_{\mathrm{j}}^{j} > k_{c,\mathrm{j}} \end{cases} \qquad [3.2]$$

where $K_{\mathrm{j}}^{j} = n_{\mathrm{j}}^{j} / \Delta x_{\mathrm{j}}$ is the density at time t_j of cell j, which is the first cell on the arc in the forward star of the origin. Cell j is characterised by the equilibrium flow-density relationship $\tilde{q}_{\mathrm{j}}(k)$. Cohorts are removed from the front of the cohort queue of the last cell and placed into a list of cohorts ready to depart the origin. The last cohort to be moved may be too large to satisfy the constraint of the local supply. If necessary, this cohort is split into two parts using bucket rounding of the packets. One part remains in the cohort queue, the other part completes the list of cohorts ready to depart the origin.

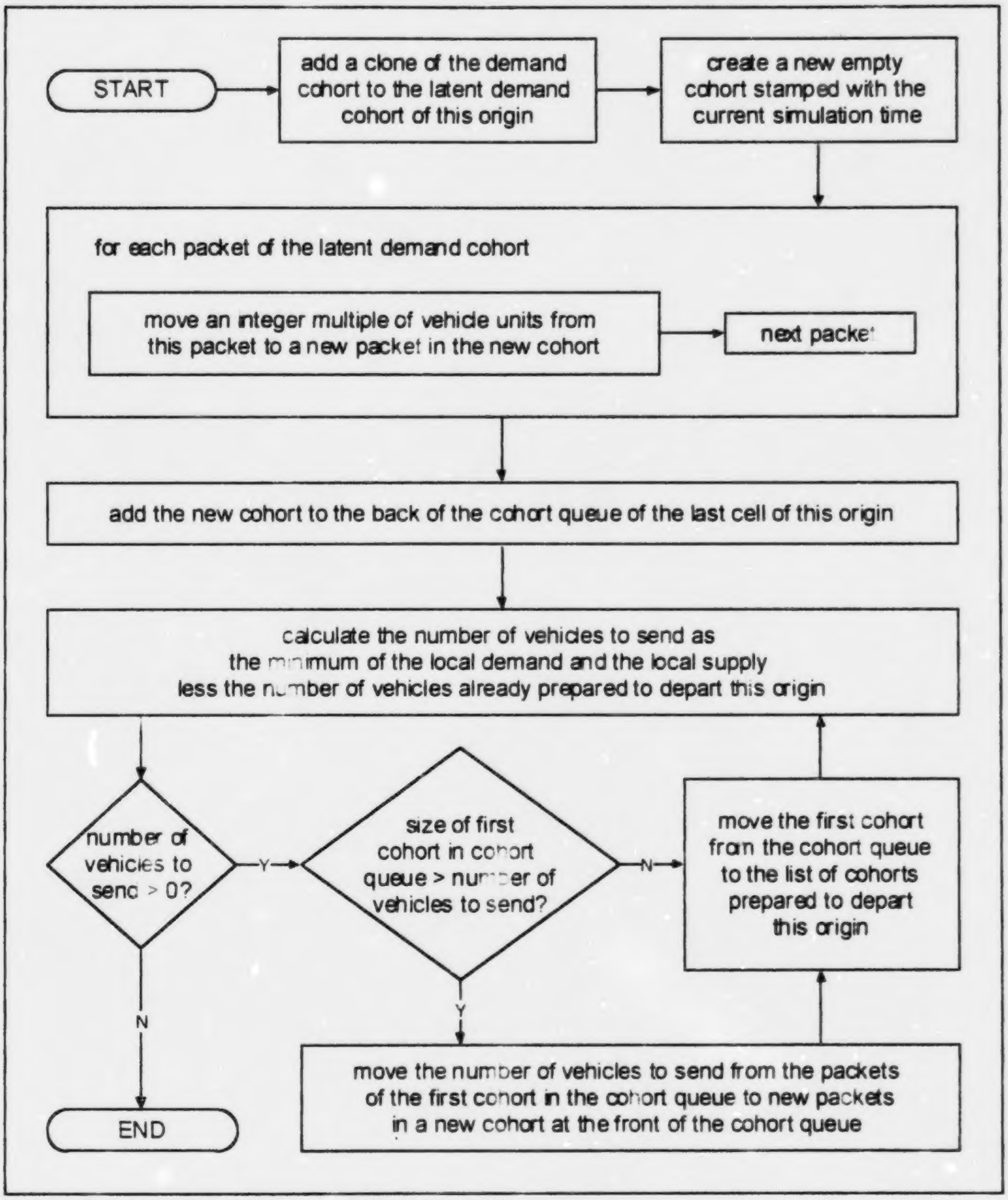

START
add a clone of the demand cohort to the latent demand cohort of this origin
create a new empty cohort stamped with the current simulation time
for each packet of the latent demand cohort
move an integer multiple of vehicle units from this packet to a new packet in the new cohort
next packet
add the new cohort to the back of the cohort queue of the last cell of this origin
calculate the number of vehicles to send as the minimum of the local demand and the local supply less the number of vehicles already prepared to depart this origin
number of vehicles to send > 0?
Y
size of first cohort in cohort queue > number of vehicles to send?
N
move the first cohort from the cohort queue to the list of cohorts prepared to depart this origin
Y
move the number of vehicles to send from the packets of the first cohort in the cohort queue to new packets in a new cohort at the front of the cohort queue
N
END

Figure 3.6 Prepare cohorts to depart an origin

A node which serves as a destination is also a special case, since the local supply is infinite. The local demand of the last cell i in the arc of the backward star is

[3.3]
$$Q^{j}_{D,i} = \begin{cases} \tilde{q}_{i}(K^{j}_{i})\Delta t & \text{if } K^{j}_{i} < k_{c,i} \\ q_{\max,i}\Delta t & \text{if } K^{j}_{i} \geq k_{c,i} \end{cases}$$

where $K^{j}_{i} = n^{j}_{i} / \Delta x_{i}$ is the density at time t_{j} of cell i. Cell i is characterised by the equilibrium flow-density relationship $\tilde{q}_{i}(k)$.

If the node is neither an origin or a destination, then cohorts must be prepared in each arc of the backward star for departure to each arc in the forward star. If the forward and backward stars have only a single arc then the node is merely a connector of two arcs, perhaps with different flow density relationships. In this case, the local supply is calculated according to [3.2], where cell j is the first arc in the arc of the forward star, and the local demand is calculated according to [3.3], where cell i is the last cell in the arc of the backward star. Cohorts are removed from the front of the cohort queue of cell i and placed in the list of cohorts ready to depart cell i, such that the total number of vehicles ready to depart cell i is the minimum of the local demand and the local supply.

If the forward and/or backward stars have multiple arcs, then the propagation of flow across the node is more complicated. If the node has multiple arcs in the backward star, but a single arc in the forward star, then flow across this node is governed by merge logic. The local supply of the merge cell, the first cell of the arc in the forward star, must be shared by the last cell of arcs in the backward star. However, if the node has multiple arcs in the forward star, but a single arc in the backward star, then the flow across the node is governed by diverge logic. The discharge capacity of the diverge cell, the last cell of the arc in the backward star, must be shared by all streams exiting the arc to different arcs in the forward star. In the basic cell-transmission model, nodes with multiple arcs in both the backward star and the forward star are not permitted. CellNetLoad overcomes this limitation by incorporating a generalised intersection logic.

Nodes with multiple arcs in the forward star or the backward star or both are called intersections. All flow across an intersection is handled by the intersection logic, which is listed in the form of an algorithm in Figure 3.7. The subscript $\mathrm{i} \in \mathrm{A}_{\mathrm{n}}^{-}$ is used interchangeably for arcs in the backward star of node n, or the last cell of any of those arcs. The subscript $\mathrm{j} \in \mathrm{A}_{\mathrm{n}}^{+}$ is used interchangeably for arcs in the forward star of node n, or the first cell of any of those arcs. The local demand $D_{\mathrm{i}} = Q^{j}_{D,\mathrm{i}} \|\mathrm{L}_{\mathrm{i}}\|$ is calculated according to [3.3]. The local supply $S_{\mathrm{j}} = Q^{j}_{S,\mathrm{j}} \|\mathrm{L}_{\mathrm{j}}\|$ is calculated according to [3.2]. The discharge capacity $Q_{\mathrm{i}} = q_{\mathrm{max,i}} \Delta t \|\mathrm{L}_{\mathrm{i}}\|$ is also stored for use in calculations. Barred variables represent remainders of that quantity. Note that the subscripts n for node and j for the time step have been dropped for convenience. Superscript m is used to denote the iteration.

Figure 3.7 Cohort FIFO algorithm
- overtaking is permitted for packets in the same cohort

```
Step 0. Initialization
        iteration $m = 0$
        vehicle movement $y_{\mathrm{ij}}^{m} = 0 \ \forall \mathrm{i} \in \mathrm{A}^{-}, \mathrm{j} \in \mathrm{A}^{-}$
        local demand remainder $\overline{D}_{\mathrm{i}}^{m} = D_{\mathrm{i}} \ \forall \mathrm{i} \in \mathrm{A}^{-}$
        discharge capacity remainder $\overline{Q}_{\mathrm{i}}^{m} = Q_{\mathrm{i}} \ \forall \mathrm{i} \in \mathrm{A}^{-}$
        local supply remainder $\overline{S}_{\mathrm{j}}^{m} = S_{\mathrm{j}} \ \forall \mathrm{j} \in \mathrm{A}^{+}$
        packet groups $n_{\mathrm{ij}\tau}^{m} = \sum_{\mathrm{d}} n_{\mathrm{ij}\tau}^{\mathrm{d}} \ \forall \mathrm{i} \in \mathrm{A}^{-}, \mathrm{j} \in \mathrm{A}^{+}, \tau \in (0, \bar{t})$
        time stamp of the front cohort $\tau_{\mathrm{i}} = 0 \ \forall \mathrm{i} \in \mathrm{A}^{-}$
Step 1. Calculate partial demand
        Step 1a. Determine the front cohort
        $\forall \mathrm{i} \in \mathrm{A}^{-}$ : while $\sum_{\mathrm{j}} n_{\mathrm{ij}\tau_{\mathrm{i}}}^{m} = 0$ and $\tau_{\mathrm{i}} < \bar{t}$ repeat $\tau_{\mathrm{i}} \leftarrow \tau_{\mathrm{i}} + \Delta t$
        Step 1b. Check that the front cohort does not exceed the local demand.
        $\forall \mathrm{i} \in \mathrm{A}^{-}$ : if $\sum_{\mathrm{j}} n_{\mathrm{ij}\tau}^{m} > \overline{D}_{\mathrm{i}}^{m}$ then split the cohort with time stamp $\tau_{\mathrm{i}}$
        into two cohorts with time stamps $\tau_{i1} = \tau_i$ and $\tau_{i2} = \tau_i + \varepsilon$
        such that $\sum_j n_{ij\tau_{i1}}^m = \overline{D}_i^m$
    Step 1c. Filter out partial demand with no local supply
        partial demand $n_{ij}^m = \begin{Bmatrix} n_{ij\tau_i}^m & \text{if } \overline{S}_j^m > 0 \\ 0 & \text{otherwise} \end{Bmatrix} \forall i \in A^-, j \in A^+$
Step 2. Calculate the number of vehicles $\Delta y_{ij}^m$ to depart cell $i$ for cell $j$
    $\Delta y_{ij}^m = \min\left(n_{ij}^m, \alpha_{ij}^m \overline{Q}_i^m, \beta_{ij}^m \overline{S}_j^m\right) \forall i \in A^-, j \in A^+$
    where $\alpha_{ij}^m = n_{ij}^m \Big/ \sum_j n_{ij}^m$, $\beta_{ij}^m = n_{ij}^m \Big/ \sum_i n_{ij}^m \quad \forall i \in A^-, j \in A^+$
Step 3. Stopping criterion
    If $\exists i \in A^-, j \in A^+ : \Delta y_{ij}^m > 0$, then continue to Step 4. Otherwise, stop.
Step 4. Update
    $y_{ij}^{m+1} = y_{ij}^m + \Delta y_{ij}^m \quad \forall i \in A^-, j \in A^+$
    $\overline{D}_i^{m+1} = D_i - \sum_j y_{ij}^{m+1}$, $\overline{Q}_i^{m+1} = Q_i - \sum_j y_{ij}^{m+1} \quad \forall i \in A^-$
    $\overline{S}_j^{m+1} = S_j - \sum_i y_{ij}^{m+1} \quad \forall j \in A^+$
    $n_{ij\tau_i}^{m+1} = n_{ij\tau_i}^m - \Delta y_{ij}^m \quad \forall i \in A^-, j \in A^+, \tau_i \in (0, \bar{t})$
    $m \leftarrow m + 1$
    Return to Step 1.
```

The partial demands n_{ij}^m are calculated in step 1 based on the occupancy in the front cohort of the cohort queue in each cell $i \in A^-$. The local demand acts as an upper bound on the total partial demand considered during one time step. If the front cohort is bigger than the remaining local demand, then the cohort is split so that the front cohort is equal to the remaining local demand. The rest of the cohort is placed behind the front cohort in the cohort queue. Only packets with remaining local supply are included in the partial demand.

In step 2, the flow, which is to be sent across the intersection during this time step, is calculated for each stream as the minimum of the partial demand, the share of the discharge capacity and the share of the local supply. The discharge capacity and the local supply are distributed to the competing streams according to the partial demands. This distribution strategy respects the constraint of the local demand on each cell $i \in A^-$ and the constraint of the local supply on each cell $j \in A^+$. In addition, any unused residual local supply is reallocated to other arcs in the backward star in subsequent iterations, and any unused residual local demand is reallocated to other arcs in the forward star in subsequent iterations.

In step 3, the algorithm stops when no more vehicles can be sent across the intersection, because at least one of the cohort queue, local demand, discharge capacity, or local supply is exhausted for each stream. If a packet in the front cohort is headed to an arc with no remaining local supply, then it is blocked from exiting that arc. Other packets in that same cohort, which are headed to a different arc, may be moved out of the cohort queue and into the cohort list of vehicles ready to depart that last cell. No other cohorts on that arc will able to exit the arc during this time step. Implicitly, first-in-first-out discipline is enforced at the cohort level, but packets in the same cohort may overtake one another.

The variables are updated in step 4 as the packets, or portions of packets, are removed from the cohort queue of the last cell $i \in A^-$. These vehicles are stored temporarily in the same last cell in an array of cohort lists, one list for each arc in the forward star.

Step 1c serves to reduce the unused residual local supply and local demand. Packets are only considered as partial demand in Step 1b if they are headed to an arc in the forward star with remaining local supply.

$$[3.4] \qquad n_{ij}^{m} = \begin{cases} n_{ij\tau_i}^{d} & \text{if } \overline{S}_j > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^{-}, j \in A^{+}$$

This adjustment avoids distributing a portion of the discharge capacity to a stream that cannot use it. The same is already accomplished for the distribution of the local supply by ensuring that the partial demands never exceed the remaining local demand in step 1b. This concludes the discussion on the intersection model.

Now that the cohorts, which are prepared to exit each origin and each arc have been placed in the appropriate cohort list of the last cell, the task of preparing flow to depart other cells is undertaken. The list of cells for each arc is traversed. In each cell, except for the last cell, the local demand is calculated from [3.3], and the local supply of the next cell is calculated from [3.2]. The cohorts are removed from the front of the cohort queue in each cell and placed in a cohort list in the same cell. Enough cohorts, or fraction of a cohort, are moved to meet the lower of the local demand or the local supply. As the list of cells is traversed, the total occupancy of each cell is added to the arc occupancy statistic.

The flow must now be sent from every cell. The cell list of each arc is traversed. The cohort list is emptied into the rear of the cohort queue of the next cell in the list. If a cohort had been split in a given cell in a previous time step, then it will be merged again in the next cell. This reduces the total number of cohorts, which improves the simulation execution time and reduces the required memory, without affecting first-in-first-out discipline.

The last cell is handled differently. Firstly, the size of all cohorts in the cohort queue is added to the arc occupancy statistics. Next, cohorts in each cohort list are merged into a single cohort. Before merging the cohorts in the cohort lists, the arc travel time statistic must be updated. The arc travel time of this cohort is derived from the current simulation time minus the cohort time stamp. The product of the size of the cohort and the arc travel time is added to the interval of the arc travel time statistic that matches the time stamp of the cohort. This ensures that a weighted average may be obtained for the arc travel time at any given *entrance* time. Other statistics are also recorded. The size of the merged cohort is added to the arc exiting vehicles statistic, and to the entering vehicles statistic of the downstream arc. Because this merged cohort is theoretically moving to the next arc during the entire time step, half of the merged cohort size is added to the arc occupancy statistic, and the other half is added to the occupancy statistic of the downstream arc. Finally, the merged cohort is placed at the rear of the cohort queue of the first cell of the downstream arc. The merged cohort will be merged with other cohorts already in that cohort queue if they share the same time stamp. This ensures that the order in which the arcs are processed for sending flow does not affect the simulation results.

The cohort list in the last cell of each origin is processed in the same way. The cohort list is merged into a single cohort. Statistics are gathered for the origin occupancy, outflow, and travel time (in the case of spillback) and the downstream arc inflow and occupancy. Then the merged cohort is transferred to the rear of the cohort queue of the downstream arc.

The interdependency of the outflows across the intersections reduces the degree to which the cell-transmission model can benefit from a parallel implementation. The preparation of cohorts to depart each ordinary cell on an arc can be performed in parallel. The last cells discharging into a particular node cannot have their cohorts prepared for departure independently, since the outflow depends on the occupancy of all the last cells on arcs in the backward star of the node and the occupancy of all the first cells in the arcs of the forward star of the node. Though each node may be processed in parallel. The actual movement of cohorts from one cell to the next can be performed in parallel for all cells, including the last cells of origins and arcs.

The last task of the clock tick event is to schedule another clock tick event for the current simulation time plus one time step. Other events scheduled for that same time take precedence, so that all of the network data structures are updated before calculating flow movements. An event which is processed between clock ticks has the same impact on the propagation of flow no matter when it was scheduled during the clock tick. The resolution of changes in routing, demand, signal timing; and incidents are implicitly limited in terms of their occurrence. In effect, all events are modelled as if they occurred just before the next clock tick.

3.5.2.6 Output statistics event

The output statistics event simply appends the current simulation time and the statistics of each origin and each arc to a text file. The statistics are then reinitialised. This reduces the memory requirements of CellNetLoad significantly. The space required to store statistics increases with the number of origins and arcs, with the simulation duration, and with the amount of congestion. Since the amount of congestion may be unknown before the simulation is run, this precaution reduces the chance that the computer memory will be exhausted during the simulation. The last task of the output statistics is to schedule another output statistics event for the current simulation time plus the output statistics interval. The output statistics event takes precedence over all events scheduled at that time. This ensures that the output statistics are emptied before a clock tick event is processed, should they share the same schedule time.

3.5.2.7 Simulation terminate event

The simulation terminate event stops the processing of the event queue and initiates the post-processing of the output data.

3.5.3 Post-processing of output

All of the data generated by the simulation is stored in a text file. As each line of this text file is read sequentially, all data is stored in five arrays. The arrays contain

- the number of entering vehicles, which is the number of vehicles which entered each arc and origin during each output interval;
- the number of exiting vehicles, which is the number of vehicles which exited each arc and origin during each output interval;
- the vehicle movements, which is the number of vehicles which exited each arc in the backward star of each node and entered each arc in the forward star of the same node during each output interval;
- the total occupancy, which is the sum of the number of vehicles present on each arc and each origin during each time step of each output interval; and
- the total travel time, which is the sum of the travel time on the arc or origin of all vehicles, which *entered* the arc or origin during that output interval.

Once the entire text file has been read, the output files are generated. One text file is created for each array with the format described earlier. The average inflow (outflow), expressed in vehicles per hour per lane, is the number of entering (exiting) vehicles divided by the duration of the output interval and the number of lanes on the arc. The average origin occupancy is reported with the average arc density. The average arc density, expressed in vehicles per kilometre per lane, is the total occupancy divided by the number of time steps per output interval, the length of the arc, and the number of lanes of the arc. The average origin occupancy is the total occupancy divided by the number of time steps per output interval. Finally, the average travel time, expressed in seconds per vehicle, is the total travel time divided by the average inflow.

3.6 Conclusions

The development of CellNetLoad was motivated by the interest in evaluating the cell-transmission model, then exploring extensions to the basic theory which may improve the simulation results. Because the development of extensions required prototyping, flexibility was always chosen over computational efficiency during development. In addition, the input-output structure and interface were designed to aid the developer in assessing modifications to the code, not for ease of use for general users.

The remainder of the thesis documents the findings of the evaluation of the basic cell-transmission model and the enhancements which were developed to improve the realism of the simulation results.

Chapter 4

Quantitative Evaluation of the Cell-transmission Model

4.1 Introduction

The cell-transmission model is a discretised approximation to the LWR hydrodynamic theory of traffic flow (Lighthill and Whitham, 1955; Richards, 1956). Compared to higher-order macroscopic models, the cell-transmission model is easily implemented in a computer program. In fact, several known simulation programs are currently in use for research, for example NETCELL (Cayford, Lin and Daganzo, 1997), and models by Ziliaskopoulos and Lee (1997), as well as Lo (1999).

The error in density along a characteristic of the cell-transmission model compared to the solution of the LWR hydrodynamic model has been quantified (Daganzo, 1995b). This error becomes zero for infinitely small discretisations of space-time These findings are based on analytical solutions of the cell-transmission model. But, in order to implement the cell-transmission model, modifications or interpretations of the theory are necessary, such as the representation of continuous flow, and respect of first-in-first-out discipline at diverges.

Furthermore, the LWR hydrodynamic model, which the cell-transmission model approximates, is itself an abstract representation of traffic flow. The fluid approximation to traffic permits the propagation of shockwaves, but it cannot model the capacity reduction and delay caused by the interactions among discrete vehicles, such as lane-changing and opposed left-turns at intersections.

The cell-transmission model is not a comprehensive solution to the dynamic network loading problem. It does not specify how to model traffic signals or incidents, nor does it specify how to route vehicles along predefined paths. General intersections are not modelled explicitly by the cell-transmission model. Instead, they are represented with a series of diverges and merges, much like a freeway interchange.

Given the above considerations, how realistic are the simulation results of an implementation of the cell-transmission model?

4.2 Evaluation strategy

The objective of this chapter is to evaluate the performance of an implementation of the cell-transmission model. The CellNetLoad program was implemented in Java as an event-based simulation. The noteworthy differences between the cell-transmission model as described in theory (Daganzo, 1994, 1995a) and this implementation are the use of discrete vehicle units and the generalised intersection model. CellNetLoad is thoroughly described in the previous chapter.

The scope of the evaluation includes the formation and dissipation of congestion due to a bottleneck. Bottlenecks are created by a lane reduction, an incident, oversaturated demand, and signalisation. Queuing, spillback, and shockwaves are examined. The effect of varying the flow demand and changes to the degree of discretisation, i.e. the time step, are considered.

A series of small networks were constructed to isolate particular network features. A linear network with a bottleneck due to a lane reduction is used to examine the congested regime, queuing, and spillback. A diverge network consists of a two-lane arterial that splits into three one-lane arterials, one of which is oversaturated. A merge network is a one-lane highway and a one-lane on-ramp, which continues as a two-lane highway with an incident in the right lane to generate congestion. The final test network is a signalised intersection of an arterial and a minor street. All arcs of each test network have the same trapezoidal equilibrium flow-density relationship, which uses parameters that match an arterial roadway. One exception is the minor street of the intersection network, which has a reduced free speed and a larger density-at-capacity. The use of small test networks proved to be effective at isolating, identifying and describing the causal factors that affect the simulation results.

In order to evaluate the cell-transmission model, a realistic benchmark was selected. All networks were simulated by CellNetLoad and by INTEGRATION, a commercially available microscopic traffic simulation program.

INTEGRATION was developed in a research environment, so it happens to be flexible in modelling hypothetical, simplified networks. The car-following model is based on a non-linear speed-headway relationship, which can be transformed into a non-linear flow-density relationship. A trapezoidal equilibrium flow-density relationship was fit to this non-linear curve for use in CellNetLoad. The levels of stochasticity can often be controlled by the user. When stochastic modelling is desired, multiple runs with different random number seeds can be averaged to reduce the "noise" in the simulation output without sacrificing time resolution. (The alternative is to aggregate the simulation output in larger time intervals.)

When modelling realistic traffic networks, INTEGRATION is generally accepted by the traffic engineering community as a reasonably accurate model, certainly relative to macroscopic models. Of course, the ideal benchmark would be a set of observations of existing networks. However, real observations are expensive, noisy, and offer no flexibility in terms of network topology, geometry, and driver behaviour. Therefore, INTEGRATION was chosen to serve as a realistic traffic laboratory, but the author acknowledges that this microscopic model is not reality.

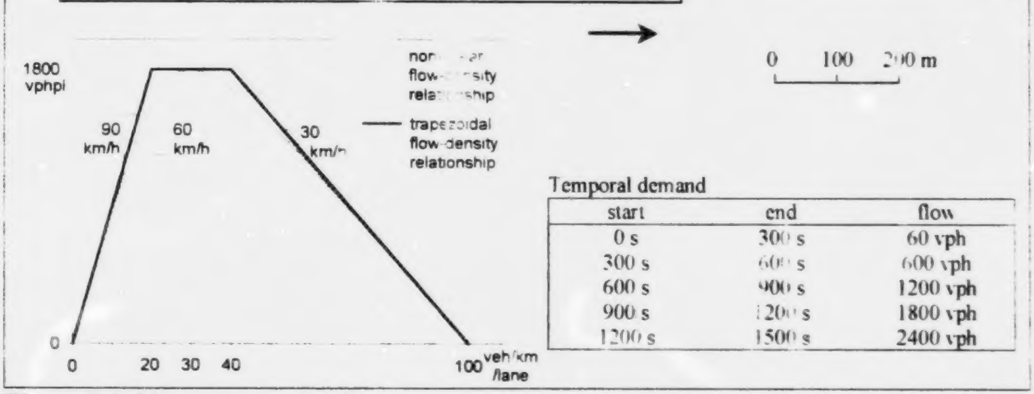
The simulation results of CellNetLoad are compared to those of INTEGRATION quantitatively and qualitatively. Large amounts of data were rendered comprehensible using a variety of graphical formats. The values of flow with respect to time are shown using line graphs for various locations in the network. Though density can be inferred from plots of cumulative inflow and outflow versus time, the formation and dissipation of congestion is difficult to visualise from flow data. Congestion is shown in space and time using 2-dimensional contour plots of density.

The intersection network presented a challenge for displaying the simulation results since the output consisted of flows of nine streams, as well as density and travel time on each approach. (There were actually 12 streams, since four approaches carried three streams each, but three streams had no flow.) Graphs plotting each of these measures against the simulation time would be cumbersome and visualising the interactions of the streams would be difficult. Instead, the simulation output was aggregated during each cycle of the signal, then these aggregate values were averaged for a number of cycles. These mean values are presented on a map of the intersection adjacent to a chart containing the same values.

The remainder of this chapter examines the simulation results for each test network. The test networks are described. The INTEGRATION and CellNetLoad simulation results are analysed and compared.

4.3 Linear network

A linear arterial roadway has an origin at position 0.0 km, and a lane drop from two lanes to a single lane 1.0 km downstream, before meeting a destination at the 1.5-km milepost, as shown in Figure 4.1. The flow-density parameters were selected to simplify analysis. The free speed is 25 m/s, which leads to integer values for the uncongested travel times on any arc that is a multiple of 25 m in length. The congested regime spans 30 to 100 veh/km/lane. The trapezoidal flow-density relationship provides capacity flows for densities ranging from exactly 20 to 40 veh/km/lane. The resulting non-linear and trapezoidal equilibrium flow-density relationships, displayed in Figure 4.1, describe a high-speed arterial roadway, though the jam density is 10 to 30 percent smaller than would be expected in practice.. The flow and density were scaled up by a factor of two for the 2-lane section of roadway. The demand was increased in 5-minute intervals from 60 vph through 2400 vph. The lane drop creates a bottleneck with capacity 1800 vph. Congestion is expected to build up upstream of the lane drop.


o
d
0
100
200 m
1800 vphpi
90 km/h
60 km/h
30 km/h
non-linear flow-density relationship
trapezoidal flow-density relationship
0
0
20
30
40
100
veh/km /lane

Temporal demand

start	end	flow
0 s	300 s	60 vph
300 s	600 s	600 vph
600 s	900 s	1200 vph
900 s	1200 s	1800 vph
1200 s	1500 s	2400 vph

Figure 4.1 Linear test network

The location of the shockwave was difficult to observe from average arc density of long arcs. So, the arcs were reduced in size, such that each arc consisted of one cell. Time steps of 20 s, 10 s, 5 s, 2 s, and 1 s were used, resulting in arc lengths of 0.5 km, 0.25 km, 0.125 km, 0.05 km, and 0.025 km, respectively. Note that the 1.0-km and 0.5-km roadway sections are evenly divisible by all arc lengths used, so there was no error in the representation of the roadway length. In order to provide adequate resolution, the collection interval was chosen to be 5 s, though it was limited to the size of the time step in the cell-transmission model.

4.3.1 INTEGRATION simulation results

The statistics were compiled over the collection intervals using arc entrance and exit times for each vehicle. The average arc inflow (outflow) is simply the number of vehicles which enter (exit) the arc during the collection interval. The average arc density is calculated by summing the travel time spent by each vehicle on the arc during the collection interval, then dividing the sum by the collection interval length. The average arc travel time is calculated by summing the arc travel times of all vehicles which *entered* the arc during the collection period, then dividing the sum by the average arc inflow.

The density contour plot (Figure 4.2) was generated by interpolating between the average density data points, expressed in vehicles per kilometre per lane. Higher density is shown with darker shading. The data points were located at the middle of the arc, at the middle of the output interval. The data points on the first arc and last arc were copied to the origin and destination. The distance between tick marks on the position axis corresponds to the arc length.

The vertical stripes in Figure 4.2 were caused by the successive increases in demand. The left-most stripe has a 30-vphpl (vehicles per hour per lane) flow, and a density of 0.33 veh/km/lane. The vehicles were travelling at 90.0 km/h when they approached the lane drop. The drivers in the right lane moved to the left, doubling the flow to 60 vphpl and the density to 0.67 veh/km/lane, and slightly reducing the speed. These density values are actually average density values, since a discrete vehicle was either present in a 25-m arc or not. In fact, close examination of the top left area of Figure 4.2 shows actual vehicle trajectories in thin grey lines. These lines are equally spaced because the departure headways were specified to be constant.

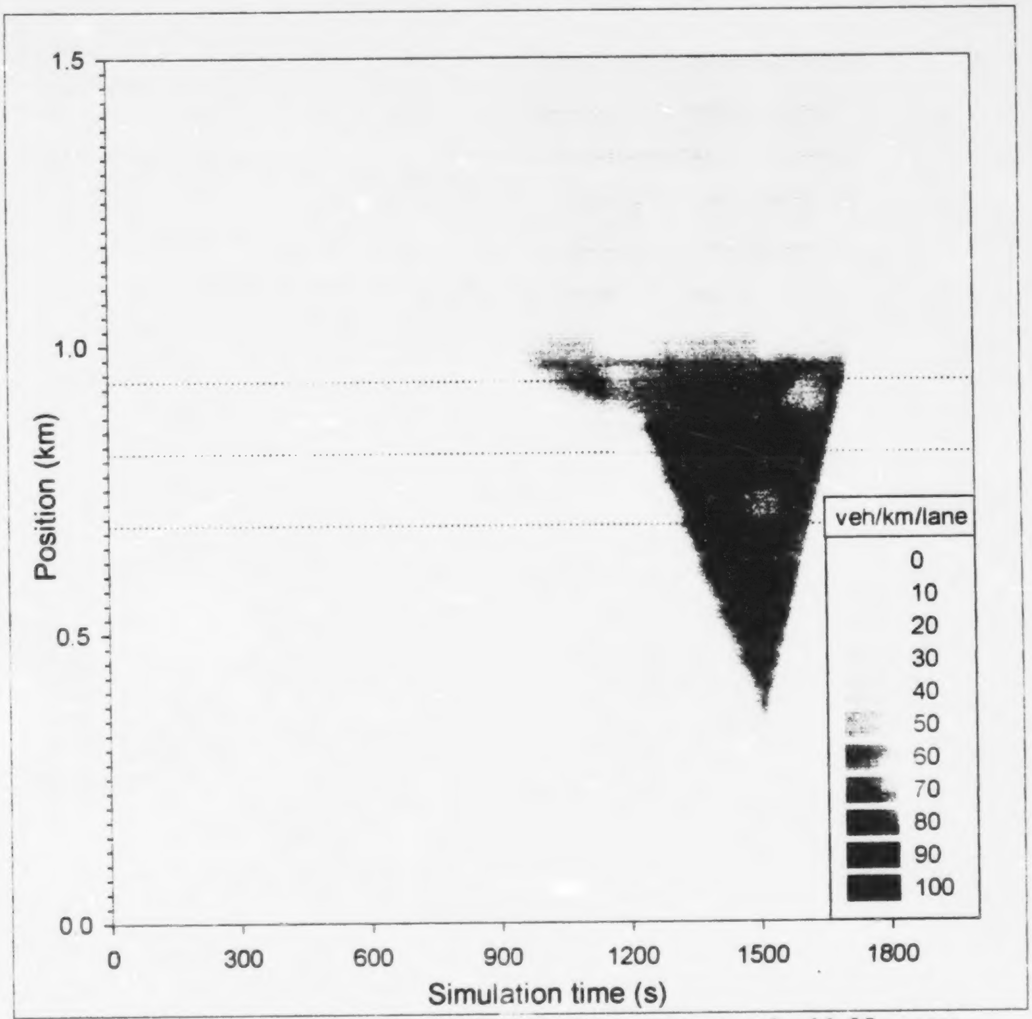


Figure 4.2 Density from INTEGRATION on the linear test network with 25-m arcs

At 300 s, the demand flow was increased to 600 vph, resulting in a flow of 300 vphpl and a density of 3.37 veh/km/lane travelling at 89.1 km/h before merging into a single lane. The first few vehicles, which departed after 300 s, were able to join the downstream flow state at a lower density, creating a fan. Because the difference in speed between the two flow states was small, the fan spread very little. The speed of the boundary was slightly less than the free speed of 90 km/h.

At 600 s, the demand flow increased to 1200 vph. The fan emanating from the origin at this time travelled at 87.0 km/h, but slowed to 81.3 km/h after the merge. These fan speeds are found by dividing the change in flow by the change in density, according to the Rankine-Huginot jump condition [2.13], assuming no spreading of the fan. The term "fan" is derived from the radiating characteristics, which are lines of equal density found on space-time plots.

At 900 s, the demand flow was increased to 1800 vph, which is the capacity of the bottleneck at the lane drop. The first vehicles arrived at the lane-drop at 940 s, found gaps in the left lane aggressively and forced vehicles already in the left lane to brake. Subsequent arrivals had more difficulty finding gaps. A queue built up, first in the right lane, then in both lanes as the queued vehicles forced their way into the left lane. As a result of this lane changing, a net capacity reduction was experienced at the lane-drop.

The demand and observed flow at the lane drop are shown in Figure 4.3. The thick solid line is the demand. The observed flow at the lane-drop is shown as thin lines for the networks with 25-m, 50-m, and 125-m arcs. The flow was aggregated in 1-minute intervals to reduce the volatility due to stochastic gap acceptance. Note the lag of the flow lines after the demand line due to the travel time to the lane drop, which was 40 seconds to 60 seconds depending on the flow rate. After reaching the lane-drop, the observed flow was approximately equal to the demand until 960 s. The gap between the demand and the observed flow is the capacity reduction due to lane changing.

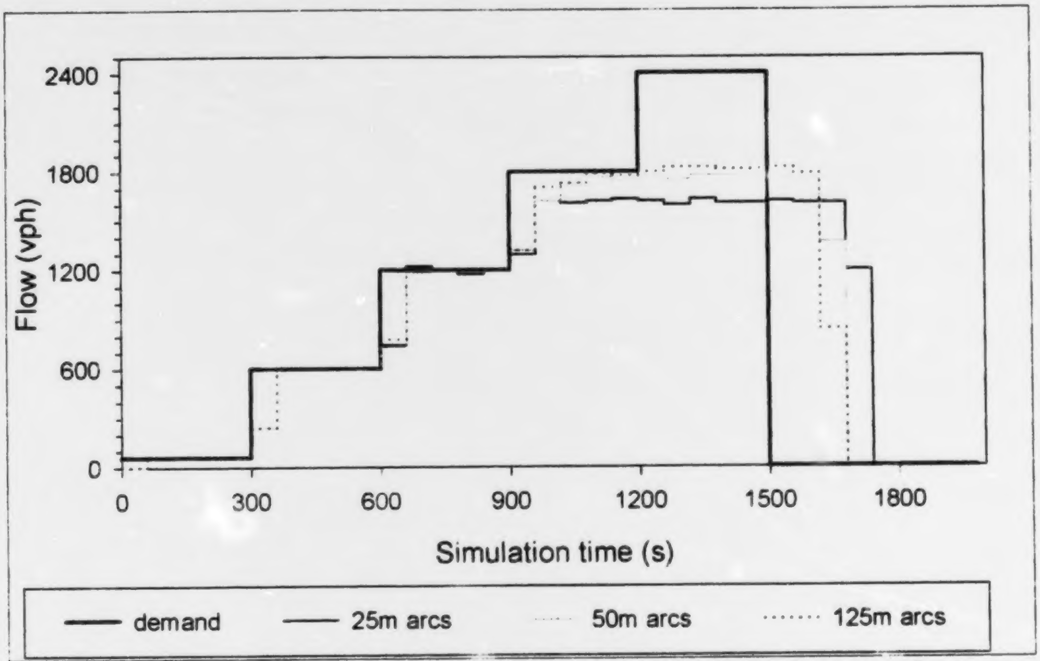


Figure 4.3 Flow from INTEGRATION at the lane drop of the linear test network

It was found that the arc length affected the size of the capacity reduction. The shorter arcs were found to cause a greater reduction in capacity. Though this effect was unforeseen, it is consistent with the lane-changing logic of INTEGRATION. Lane changing allows vehicles to briefly occupy two lanes, resulting in a capacity reduction. Mandatory lane changes occur within a window defined by a soft wall and a hard wall. A soft wall is the location in a lane at which a driver first realises that a lane change must be made in order to continue on the known path. The hard wall represents the location farthest downstream in the same lane at which the lane change manoeuvre must be made. A vehicle will slow down as it approaches the hard wall, blocking flow in that lane until it finds a suitable gap in the adjacent lane. Drivers become increasingly aggressive in accepting gaps in the adjacent lane as they approach the hard wall. Aggressive lane-changes can force the following vehicle to brake, thereby incurring some delay.

Because a vehicle is made aware of the geometry on the current arc and the downstream arc only, the window between the soft wall and hard wall was compressed for the network with 25-m arcs.

All lane changes occurred in the 50 m upstream of the lane drop, which caused drivers to be more aggressive. Longer arcs allowed the simulated drivers more time to find a suitable gap and to brake more gradually. The average capacity of the 25-m arc upstream of the lane-drop was 1616 vph in the left lane only. The right lane was dropped and had no outflow. The network with 50-m arcs had a smaller capacity reduction, yielding an average capacity of 1770 vph. Networks with longer arcs experienced no capacity reduction. Apparently, one should code an INTEGRATION network such that no more than 2 arcs span the 100-m upstream of a mandatory lane change, assuming a free speed of 90 km/h.

The choice of the benchmark is complicated by the effect of the capacity reduction predicted by INTEGRATION. In reality, some capacity reduction due to lane-changing is plausible, but INTERATION only suggests a range of possible capacity reductions. For the purpose of this study, the benchmark is assumed to be the network with 50-m arcs, since the capacity reduction was present, but the two-arc, 100-m long window for lane changing was sufficient to avoid exaggerating the capacity reduction. Networks with longer arcs lacked the resolution necessary to create useful density contour plots. Wherever possible, the cell-transmission model will be evaluated against the range of capacity reductions suggested by INTEGRATION.

Figure 4.4 looks similar to Figure 4.2, except that the minor tick marks on the position axis are spaced more widely to represent the locations of the 50-m arcs. There is some congestion as the 1800-vph demand from 900 to 1200 s reaches the 100 m upstream of the lane drop. However, there was minimal queue growth from 960 s to 1260 s, which is shown as a dark triangle in Figure 4.2. There was less queuing overall in Figure 4.4. Also, it can be observed that the arc directly upstream of the lane drop never became congested, because most lane changes took place farther upstream, leaving the right lane of the last arc empty. This feature was difficult to see on Figure 4.2, because the arc length is shorter.

The density plot is useful for performing a shockwave analysis. According to the entropy condition, a shockwave forms at any density discontinuity such that the density increases in the direction of flow. Such a situation occurs both at the lane drop and as the queue builds at the lane drop. The former is a stationary shockwave, which is caused by the lane discontinuity. The latter shockwave is the rear of the queue.

Several points in the space-time plane were carefully chosen. The queue appeared to begin at 942.5 s at 0.975 km. The queue grew slowly until 1237.5 s, when it reached position 0.925 km. The queue peaked at 1517.5 s at position 0.475 km. Clearance of the queue ended at time 1657.5. The shockwave position is shown on Figure 4.4 as the white line which joins these points. The slope of this line is the shockwave speed. The shockwave moves slowly at first, moving backward at 0.61 km/h. It travels at 5.28 km/h as the oversaturation demand meets the queue. The rear of the queue clears at 12.85 km/h.

The shockwave speed can also be calculated from the known flow states upstream and downstream of the shockwave. As mentioned earlier, the shockwave speed is the change in flow across the discontinuity, divided by the change in density. The conditions upstream of the shockwave changed as the demand changed. From 900 s to 1200 s, a demand flow of 900 vphpl at a density of 10.41 veh/km/lane met the shockwave close to the merge. The demand flow of 1200 vphpl was at a density of 14.34 veh/km/lane, according to the equilibrium flow-density curve shown in Figure 4.4. As the demand ended, the density became zero upstream of the shockwave. Each of the three uncongested flow states can be connected by a straight line to the congested flow state of approximately 900 vphpl and 70 veh/km/lane. The slopes of these lines closely match the slope of the white line in the density plot.

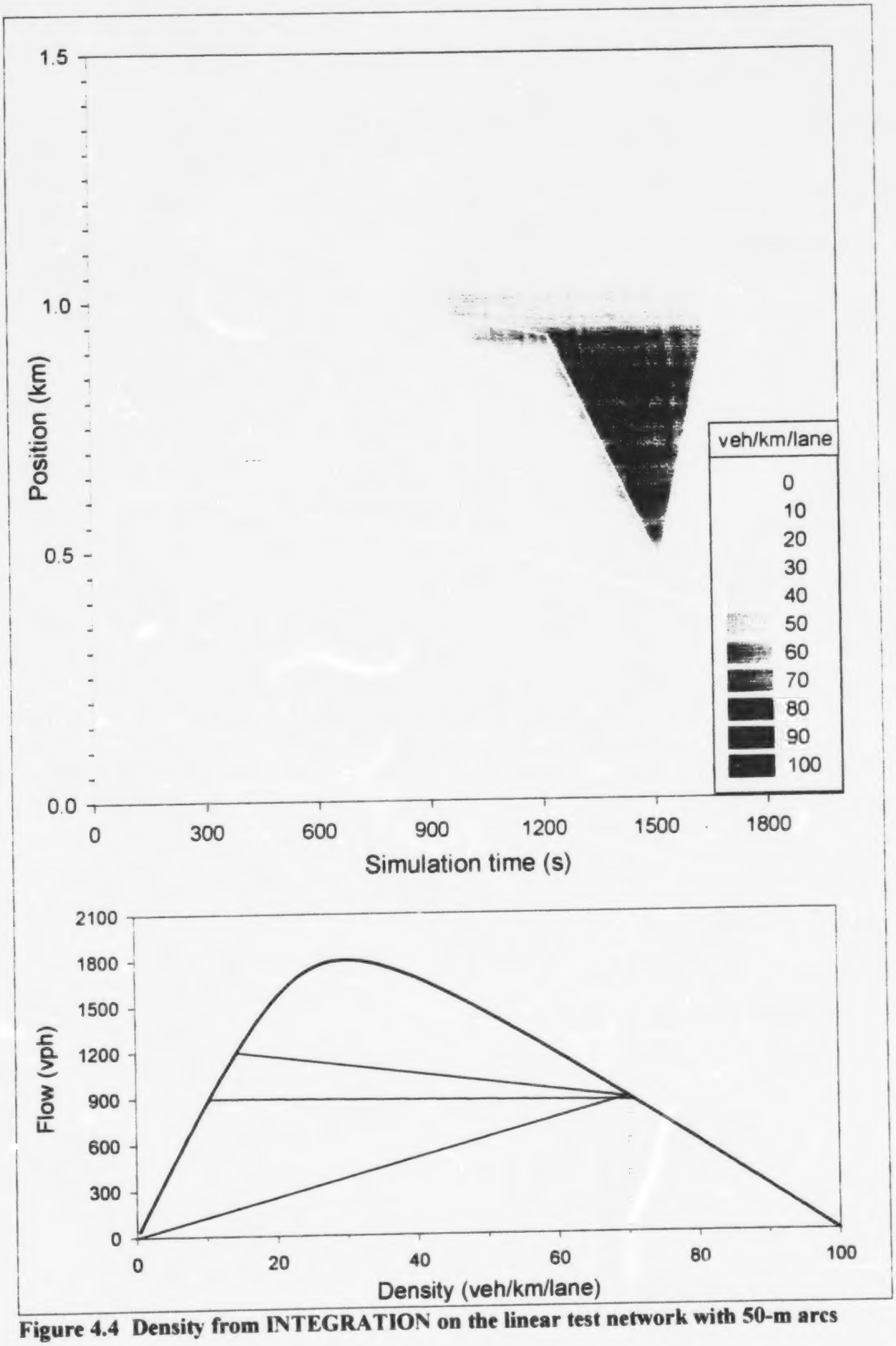


Figure 4.4 Density from INTEGRATION on the linear test network with 50-m arcs

As shown in Figure 4.3, the flow in the queue varied from 864 to 888 vphpl for the 50-m arc due to the capacity reduction described earlier. The congested flow state is not static. The average congested flow state can be determined by projecting the slope of the white line out from the uncongested flow states, and observing the intersection of the straight line with the equilibrium flow-density curve. The congested flow was found to change from 864 vphpl to 878 vph, and clear at 886 vph. These values seem to match the observed flow at the lane drop during each of the 300-s periods (Figure 4.3).

Therefore, the INTEGRATION simulation results seem to be consistent with shockwave analysis, considering the capacity reduction due to lane changing at the lane drop. The density plots are useful for shockwave analysis and relative comparisons of results. More exact density values must be displayed differently.

In order to examine the spreading of the shockwave, the density was observed on the three arcs upstream from the lane drop for the network with 125-m arcs. The density at the midpoints of these arcs is shown in Figure 4.5.

Early in the simulation, the demand flow of 60 vph generated a discrete vehicle every 60 s. The trajectory of these vehicles is shown by spikes in density of 4 veh/km/lane = 1 veh / 0.125 km 2 lanes. The 600-vph demand flow generates one vehicle every 6 s. Since the arc travel time is 5 s, there is still only one vehicle per arc. One sixth of the time the arc is empty, so the average density in any 5-s output collection interval may have a density lower than 4 veh/km/lane. At 600 s, the demand flow was increased to 1200 vph, which is equivalent to one departure each 3 seconds. Each arc has either one or two vehicles, which results in up to 8 veh/km/lane.

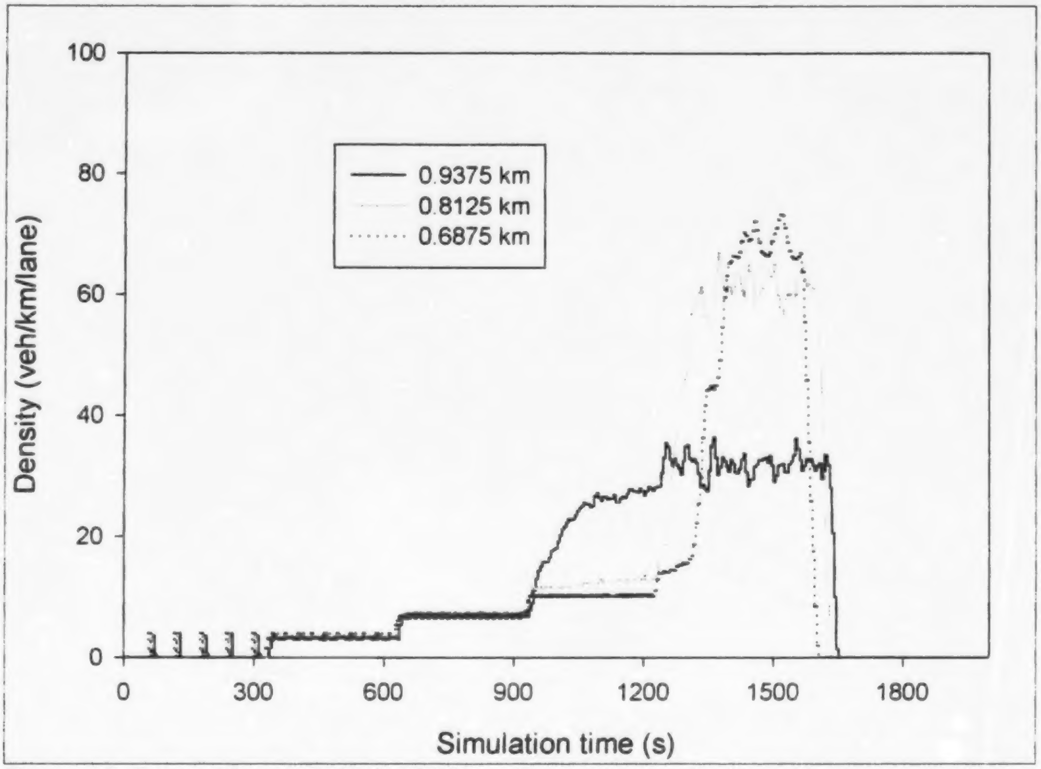


Figure 4.5 Density from INTEGRATION at fixed points on the linear test network with 125-m arcs

At 900 s the demand was equal to a single lane capacity, which is 1800 vph. The queue began on the arc just upstream of the lane drop, shown as the solid, black line in Figure 4.5. Vehicles in the right lane slowed as they approached the hard wall, looking for an acceptable gap in the left lane. If necessary, an aggressive lane change would occur that forced a vehicle in the left lane to brake. Both the slowing and braking of vehicles led to an increase in density. The queue growth slowed once the average arc density reached 25 veh/km/lane. It is important to realise that the density was not homogeneous on the 125-m arc. The left lane density varied from approximately 10 veh/km/lane at the arc entrance, which corresponds to a flow of 900 vphpl, to 30 veh/km/lane at the arc exit, which corresponds to a capacity flow of 1800 vphpl. (No capacity reduction was noted for this network with 125-m arcs.) The density in the right lane varied from 10 veh/km/lane at the arc entrance, but rose to approximately 70 veh/km/lane as the vehicles queued at the hard wall. Since vehicles do not continue past the hard wall, the density on the remainder of the arc was 0 veh/km/lane.

The oversaturated demand of 2400 vph began at 1200 s. Both lanes quickly became congested due to aggressive lane changes, but the roadway just upstream of the lane drop remained at capacity in one lane, and empty in the other. The portion of that arc which remained uncongested depended on the location of the hard wall, which varies with the aggressiveness of each downstream driver in the right lane. The queue quickly spilled back to the next upstream arc, as indicated by the solid, grey line in Figure 4.5. The density rose sharply as the shockwave traversed the arc. Note that the shockwave is spread over a certain distance since vehicle deceleration (and acceleration) is finite. The density remains at approximately 60 veh/km/lane because many lane changes occur on this arc, as gaps become available in the left lane. The third arc upstream of the lane drop experiences the same spillback shockwave, only it arrives 90 seconds later as expected, and the final density is closer to 70 veh/km/lane because the vehicles in the right lane are unaware that a mandatory lane change is necessary until they enter the next arc.

The same observations for the network with 25-m arcs are shown in Figure 4.6 to demonstrate the effect of the capacity reduction due to lane changing on the density. The midpoints of the third, eighth, and thirteenth arcs upstream of the lane drop correspond to the positions shown on Figure 4.5. These positions are shown with dotted, horizontal lines in Figure 4.2. The observations are the same for the first 300 s. The next 300 s on Figure 4.6 continues to show the trajectories of discrete vehicles, though more densely spaced, since the arc travel times are only 1 s. The density for the 1200-vph demand is the same again as Figure 4.5.

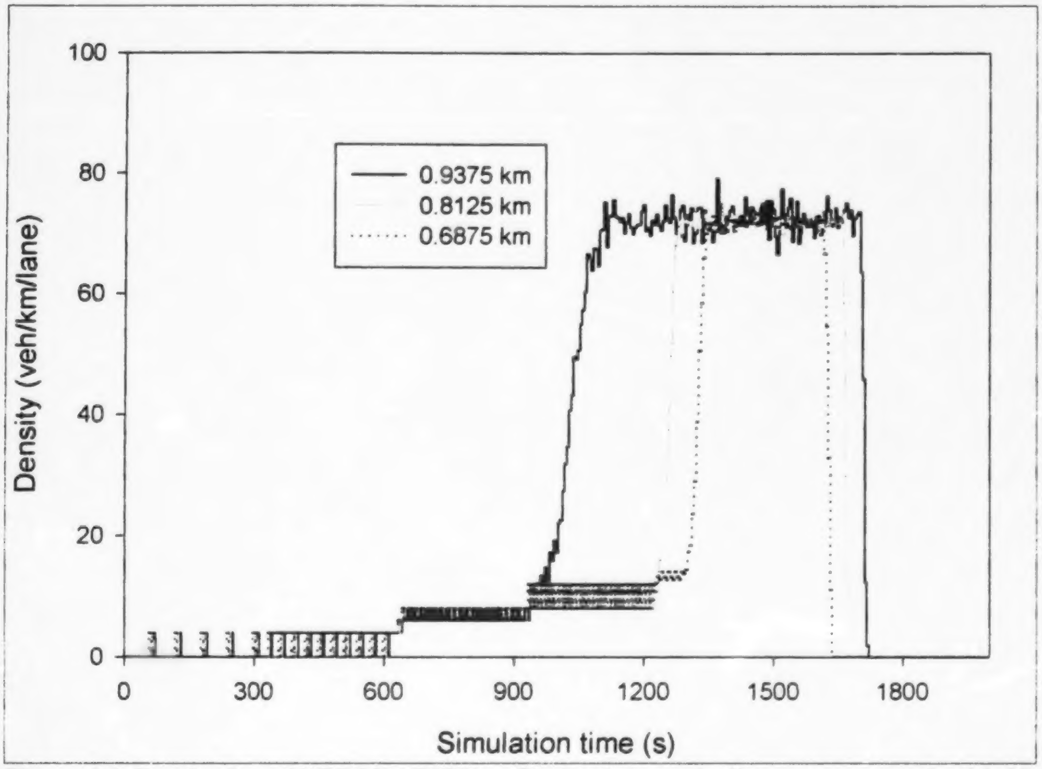


Figure 4.6 Density from INTEGRATION at fixed points on the linear test network with 25-m arcs

The main difference in the density observations occurs at 900 s, as the 1800-vph demand reached the lane drop. Due to the capacity drop, a queue formed quickly in the right lane and spilled over to the left lane. With both lanes congested, the density rose sharply to 70 veh/km/lane. The uncongested portion just upstream of the lane drop is not included in this short arc. Once the oversaturated demand departed at 1200 s, the congestion grew more quickly, crossing the eighth arc in 50 seconds, and spilling back from the eighth to the thirteenth arc in 90 seconds. This time lapse is consistent with the 5-km/h shockwave speed calculated above. The recovery shockwave takes 25 s to cross the 125 m between the arc midpoints, which is consistent with the 13 km/h shockwave speed.

The true measure of the cell-transmission model, in the context of a solution to the dynamic network loading problem, is the estimate of temporal path travel times. The benchmark path travel times are derived recursively from average arc travel times. The path travel time is equal to the arrival time at the end of the first arc, given a trip departure time, plus the path time from the beginning of the next arc, with a departure time defined by the arrival at that arc. More intuitively, a vehicle probe starts a trip on a given route at a specified departure time and uses the current estimated arc travel time to traverse each arc as it arrives at the entrance to that arc. This definition of a path travel time is consistent with the fact that average arc travel times are the mean travel time of all vehicles which *entered* the arc during the statistics collection interval.

The path travel times for three networks with different arc lengths are presented in Figure 4.7. The average arc travel times were estimated for each 5-s interval, and the path travel times were calculated with hypothetical departures every 20 s. Linear interpolation was used between the 20-s estimates of path travel time.

The path travel times were less stable for networks with shorter arcs because the average arc travel times were based on shorter, hence more variable, arc travel times. Nevertheless, the increased travel times are visible in Figure 4.7 each time the demand increased, even during uncongested conditions. The aggressive lane changing on 25-m arcs actually created temporary congestion at the lane drop, which would recover, but is reflected in the path travel times despite the small flow rates.

From 900 to 1200 s, the network with 25-m arcs experienced oversaturation, and a steadily rising path travel time. The path travel times for networks with longer arcs tended to increase, then plateau. The path travel times of all networks rose steeply during the 2100-vph demand from 1200 to 1500 s, most steeply for the network with 25-m arcs. The reduction in path travel times is consistent with the clearing of the network.

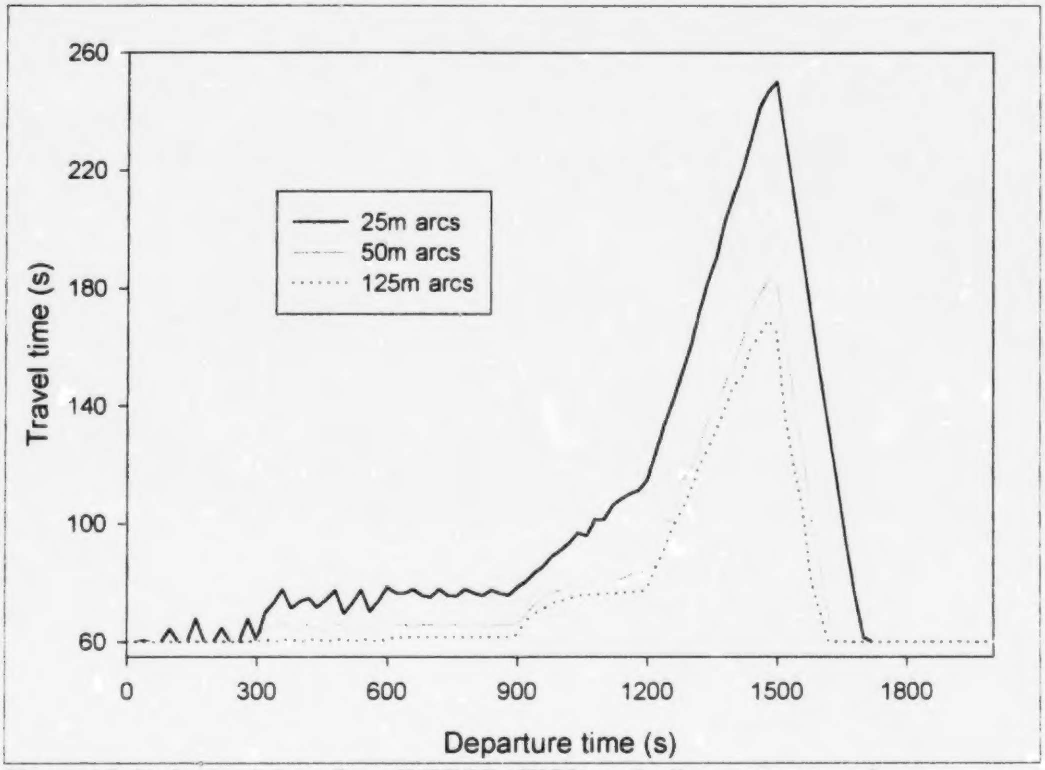


Figure 4.7 Path travel times from INTEGRATION on the linear test network

4.3.2 CellNetLoad simulation results

The network with 25-m arcs offered the best data resolution for the density contour plot, which is presented in Figure 4.8. A 1-s time step was chosen so that each arc consisted of one cell. Since the cell-transmission model does not model lanes, there is no capacity reduction due to lane changing. Therefore, the only effect of the reduced time step (and arc length) was the reduced error in approximating the LWR hydrodynamic model.

Vehicles are modelled as a continuous fluid, so the density is smooth in the top, left corner of Figure 4.8, unlike Figure 4.2 and Figure 4.4. Due to the trapezoidal equilibrium flow-density relationship, all uncongested flow travels at the free speed.

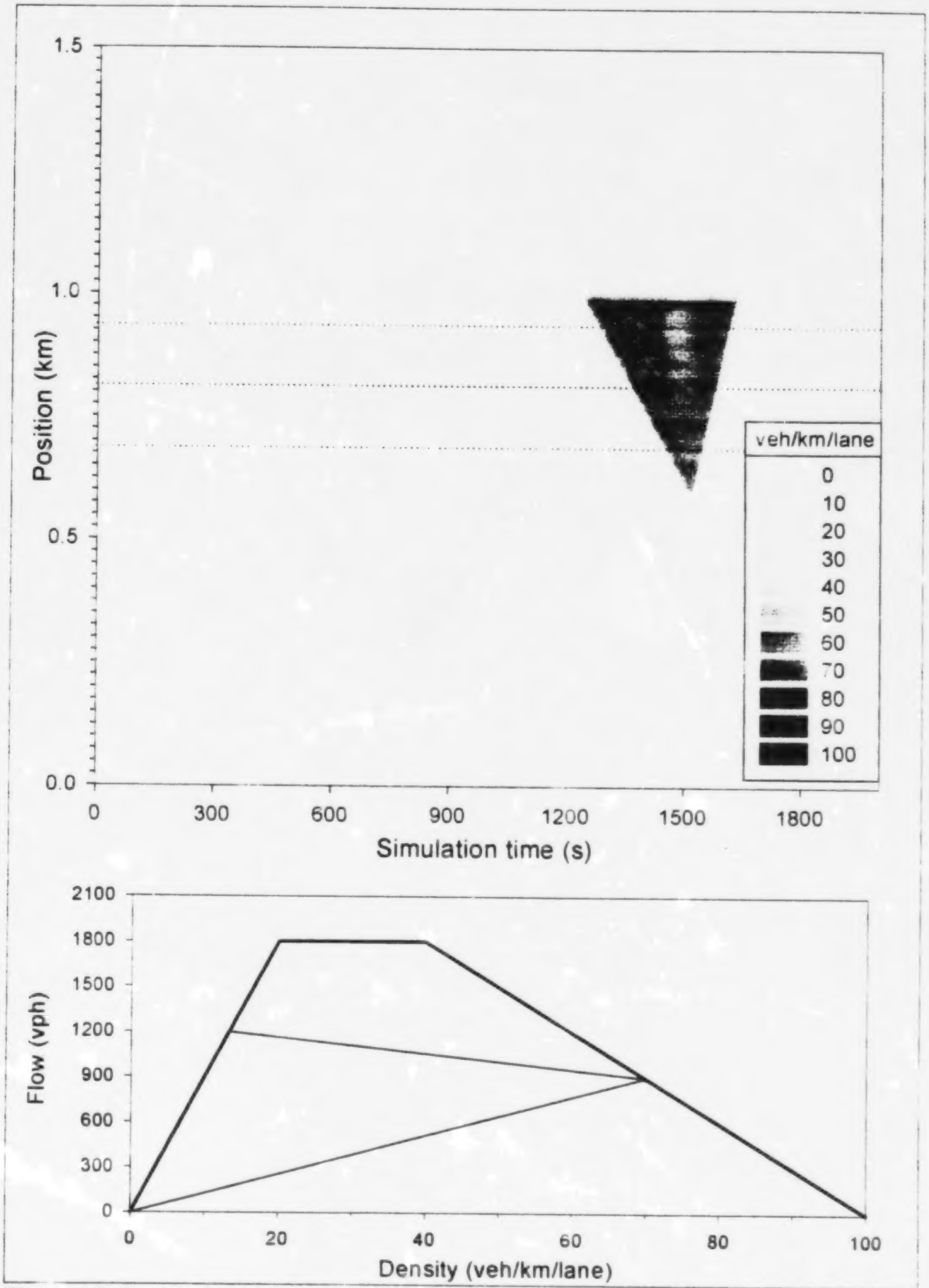


Figure 4.8 Density from CellNetLoad on the linear test network with 25-m cells (1-s time step)

The boundaries between the uncongested flow states also travelled at the free speed. The boundaries are actually fans. In the LWR hydrodynamic model, these fans do not spread, which implies that the density discontinuity remains intact as it moves downstream, behaving like a shockwave. The entropy condition (Ansorge, 1990) states that all shockwaves must have increasing density across the discontinuity in the direction of flow. Therefore, this boundary violates the entropy condition, as the density decreases across the discontinuity.

The shockwave analysis for the congestion is similar to INTEGRATION. There was no queuing present for the demand flow of 1800 vph, because the capacity at the bottleneck was exactly 1800 vph. Queuing began at 1242.5 s. A shockwave, drawn as a white line on Figure 4.8, started at 0.9875 km and moved upstream at 5.1 km/h. The queue peaked at 1522.5 s at 0.5875 km, at which point the demand ceased, and the queue cleared at 12.5 km/h. The queue was completely cleared at 1637.5 s.

These observations from the simulation agree with the analytical calculations. The flow state of the queue upstream of the lane drop was 900 vphpl, with a congested density of 70 veh/km/lane. The flow state upstream of the congestion-forming shockwave was 13.33 veh/km/lane and 1200 vphpl. The straight line joining these flow states is shown in the lower diagram of Figure 4.8, and it has a slope equal to -5.3 km/h. The recovery shockwave speed was calculated to be 12.85 km/h. The small discrepancy between the analytical shockwave speeds and the observed values in the simulation can be attributed to the 25-m, 5-s resolution of the density contour plot.

By decreasing the resolution, the simulation was made to run faster, but it yielded a lower resolution density contour plot. The cell-transmission model assumes that the density is uniform on each cell, meaning that there are density discontinuities at cell boundaries, which are visible on the density contour plot in Figure 4.9.

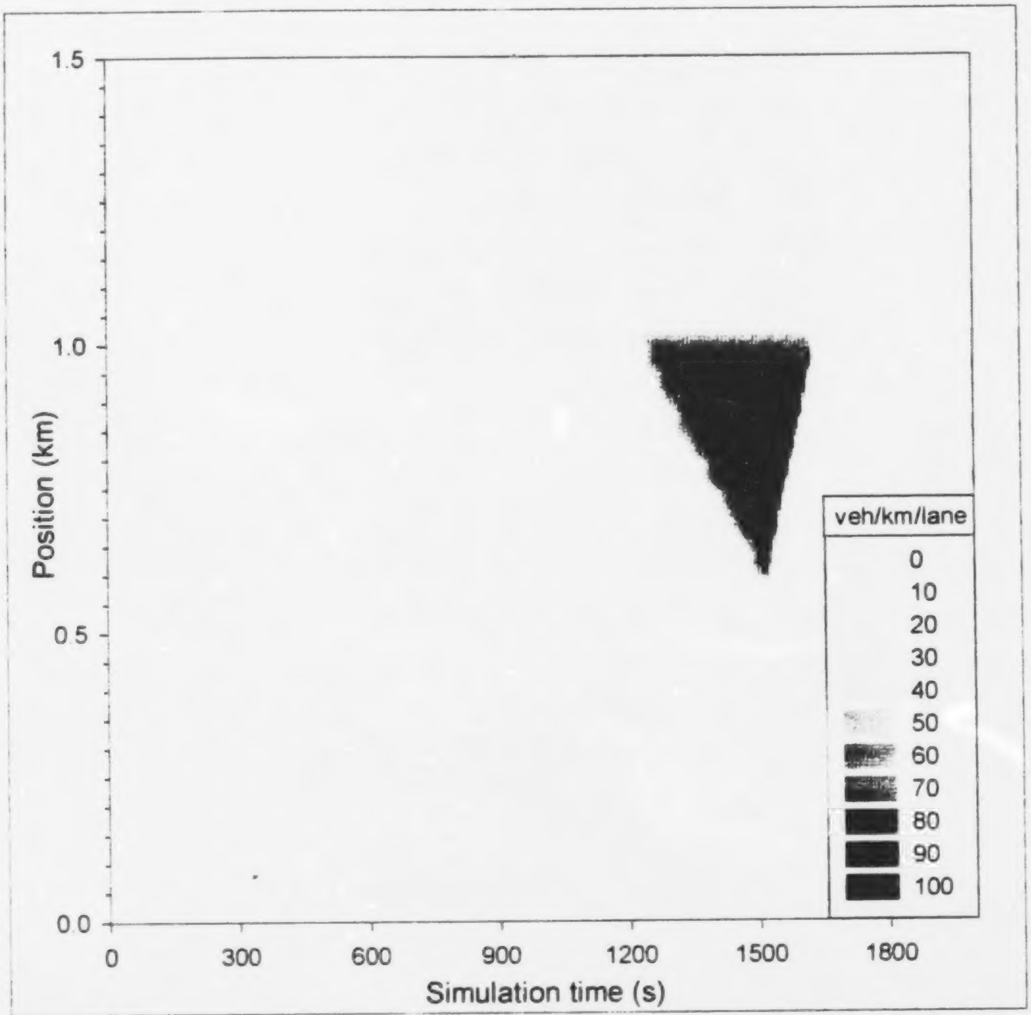


Figure 4.9 Density from CellNetLoad on the linear test network with 50-m cells (2-s time step)

It was mentioned above that shockwaves are modelled in the LWR hydrodynamic model as density discontinuities at points space-time. The cell-transmission model is an approximation to the LWR hydrodynamic model. Some spreading of the density discontinuities occurs as a result of numerical error because the position of the discontinuity within a cell is unknown. As cells are made smaller with finer discretisations, the location of the shockwave is made more precise, and less spreading of the shockwave occurs. This error is shown in Figure 4.10 and Figure 4.11.

Density during the simulation was observed at three positions, shown as dotted, horizontal lines on Figure 4.8. These density values, from the network with 25-m arcs, are shown on Figure 4.10. The fluid representation of vehicles is evident when comparing the density values for the first 300 s with Figure 4.6.

The congestion was first visible at the downstream position 0.9375 km when the density rose sharply to 70 veh/km/lane. The shockwave took 85 seconds to reach position 0.8125 km, and another 85 seconds to reach position 0.6875 km. Each time the shockwave passed a point, the density increase was spread over 2 cells (50 m) and 6 output intervals (30 s). The recovery shockwave took 1 to 2 cells (25 to 50 m) and 4 output intervals (20 s) to cross any given point.

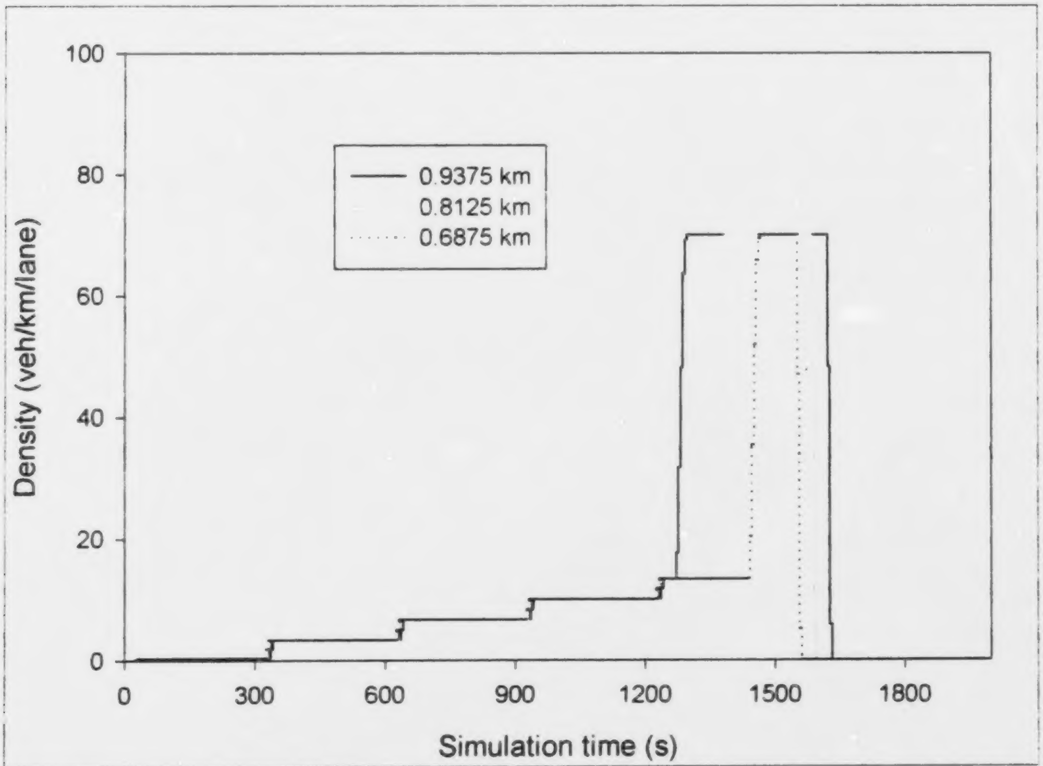


Figure 4.10 Density from CellNetLoad at fixed positions on the linear test network with 25-m cells (1-s time step)

This spreading of density discontinuities increases as the time step is increased. Figure 4.11 is a plot of density at the same positions as for Figure 4.10, but with a time step that was increased from 1 second, to 5 seconds. Consequently, the cell length increased from 25 m to 125 m. Both the backward- and forward-moving shockwaves are spread over more cells and time steps due to the coarser discretisation. As finer discretisations are used, the cell-transmission model converges to the solution of the LWR hydrodynamic model, in which these density changes occur instantaneously.

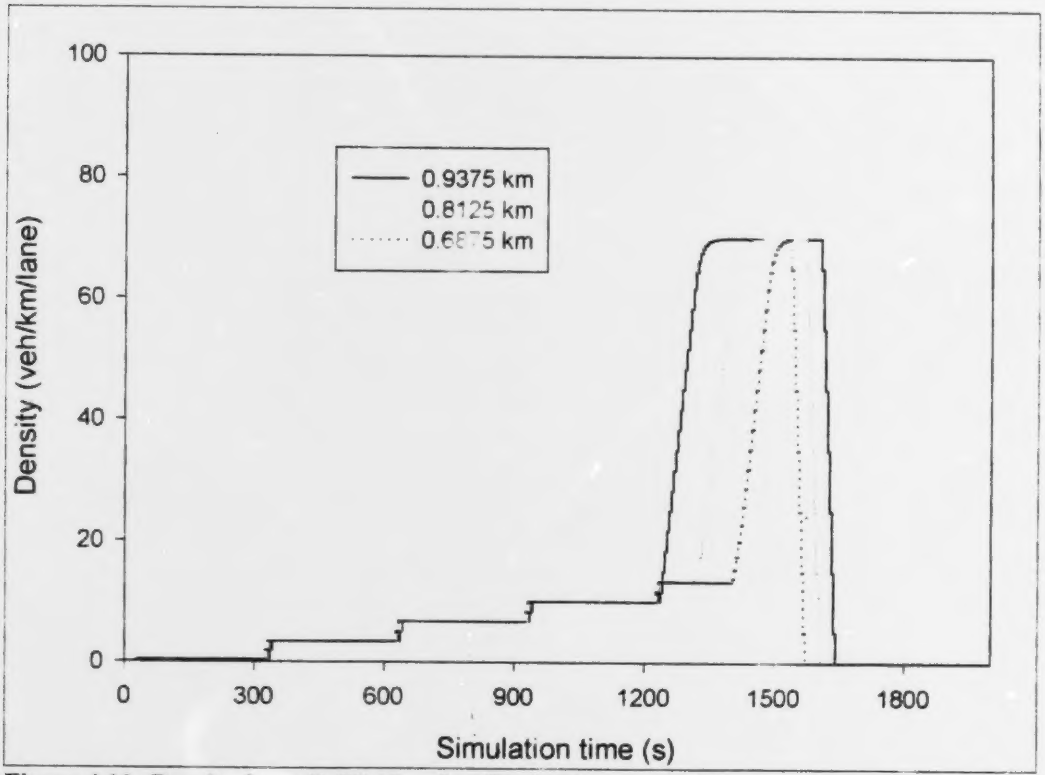


Figure 4.11 Density from CellNetLoad at fixed points on the linear test network with 125-m cells (5-s time step)

The path travel times for each time step are presented in Figure 4.12, calculated in the same fashion as for INTEGRATION. During uncongested conditions the cell-transmission model predicts that the path travel time is the travel time at the free speed. The first congestion occurs at 1200 s with the oversaturation demand. The path travel time rises steeply for all the simulations, but the simulation with a 20-s time step had a wavy estimated path travel time due to the lower data resolution.

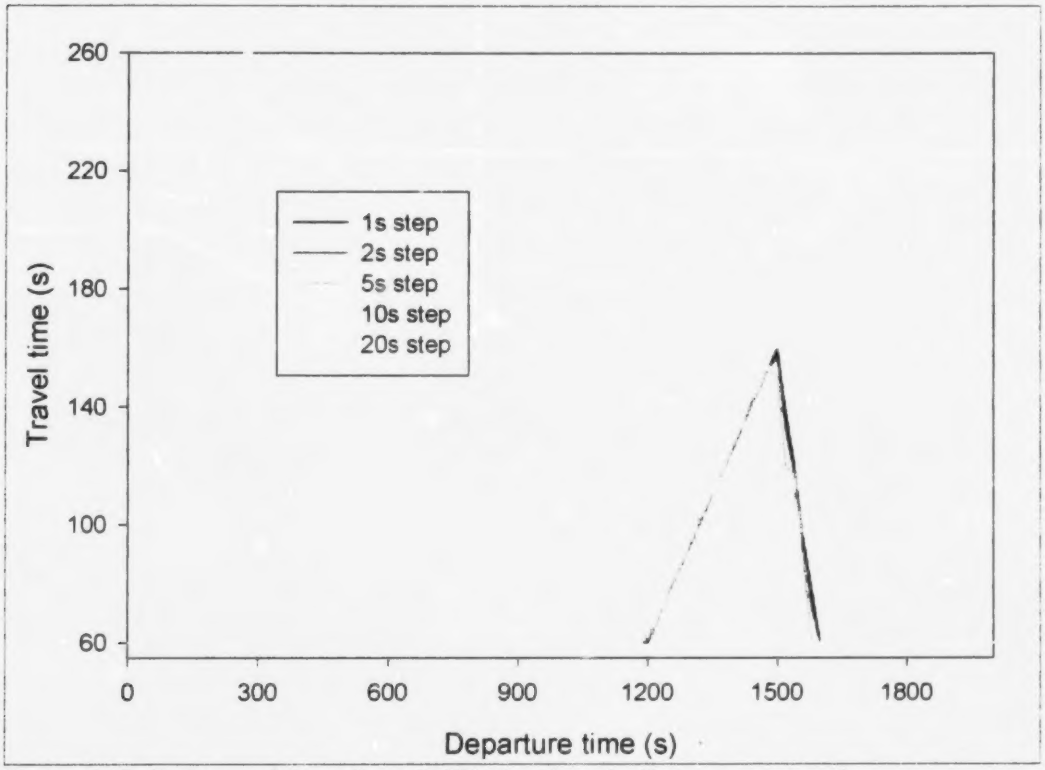


Figure 4.12 Path travel times from CellNetLoad on the linear test network

The congestion seems to clear more quickly for simulations with longer cells, but this is merely a side-effect of the use of vehicle probes to estimate arc travel times. The queue was contained on the 400 m upstream of the lane drop. The simulation with a 20-s time step had 500-m arcs with one cell per arc, so the queue was entirely contained on the one cell upstream of the lane drop. Since no vehicles entered that arc/cell after 1520 s, the arc travel time estimate was the travel time at the free speed. Had a vehicle departed after 1500 s and joined the queue after 1520 s, then it would have been used as a probe in the estimate of the arc travel time. This same effect occurred in the simulation with a 10-s time step and 250-m arcs/cells, but to a lesser degree because vehicles in the queue were delayed in entering the cell upstream of the lane drop. These delayed vehicles acted as probes which included part of the queuing delay. Therefore, with no demand, the arc travel times are unreliable because of a lack of probes, though the effect is diminished for smaller time steps.

CellNetLoad is consistent with analytical shockwave analysis. For the simulation of this linear network it differed from INTEGRATION mainly as a result of the capacity reduction due to lane changing. Because the flow was equal to capacity in CellNetLoad, congestion began only for an oversaturated demand flow. It also occurred immediately upstream of the lane drop, whereas INTEGRATION showed that the vehicles changed lanes at least 50 m upstream of the lane drop. The rates of congestion propagation and clearance of the queue were similar in both simulations.

There is no mechanism in the cell-transmission model to model lane changing explicitly because of the fluid representation of the traffic. But, if the capacity reduction could have been predicted somehow, then the specified capacity of the cells could have been explicitly reduced in the input files to account for the effect of lane changing. The derivation of methods to predict the impact of lane-changing on capacity is beyond the scope of this work.

The capacity reduction predicted by INTEGRATION was used to determine the specified reduced capacity of the arc immediately upstream of the lane drop in the CellNetLoad simulations. The reduced capacity of 1616 vph was distributed on the two lanes of the 25-m arc, yielding a reduced capacity of 808 vphpl for the CellNetLoad simulation with a 1-s time step and 25-m cells. Figure 4.13 shows the resulting congestion in the CellNetLoad simulation, which forms and clears similarly to INTEGRATION, as shown in Figure 4.2.

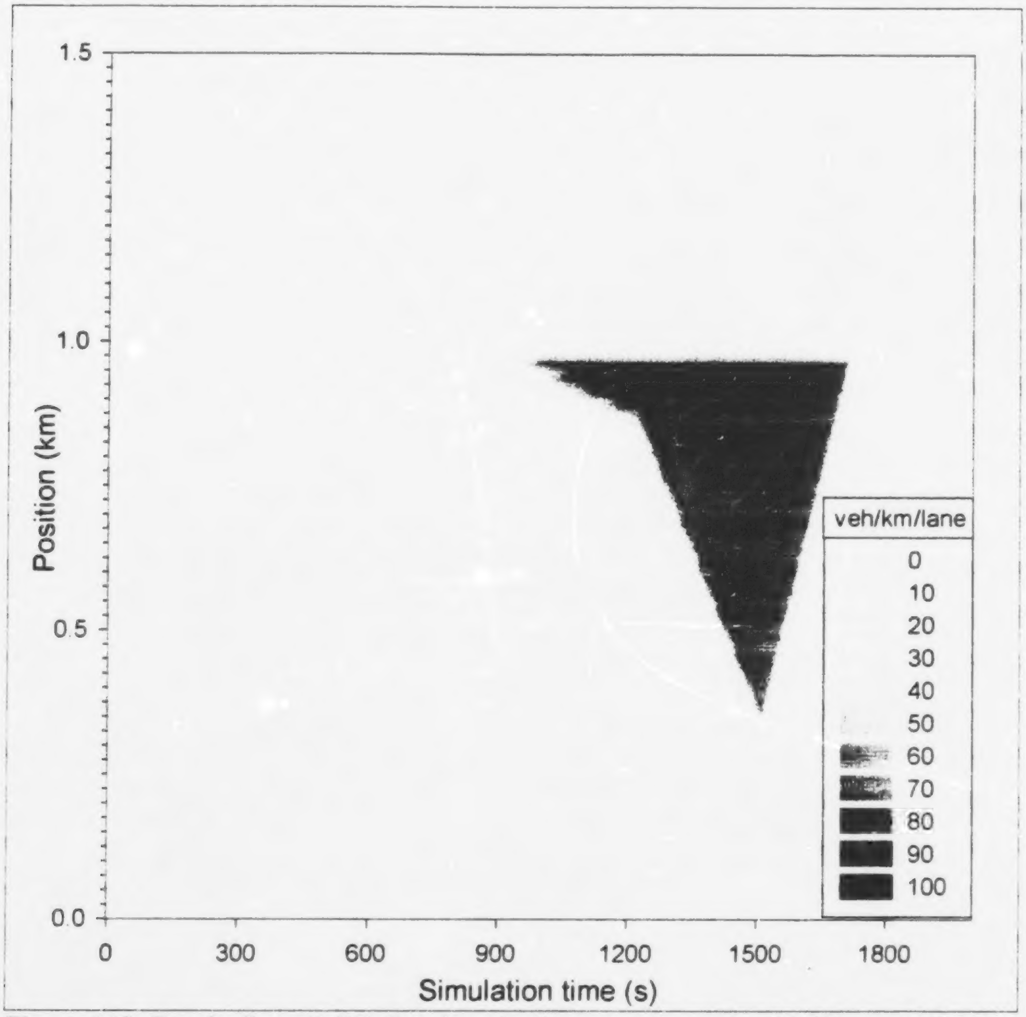


Figure 4.13 Density from CellNetLoad on the linear test network with 25-m arcs (1-s time step) and a reduced capacity at the lane-drop

The reduced capacity of 1770 vph, which was predicted by INTEGRATION, was distributed on the two lanes of the 50-m arc, yielding a reduced capacity of 885 vphpl for the cell immediately upstream of the lane drop in the CellNetLoad simulation with a 2-s time step and 50-m cells. Figure 4.14 shows the resulting congestion in the CellNetLoad simulation, which forms and clears similarly to INTEGRATION, as shown in Figure 4.4.

No capacity reduction was found to occur for networks with longer arcs.

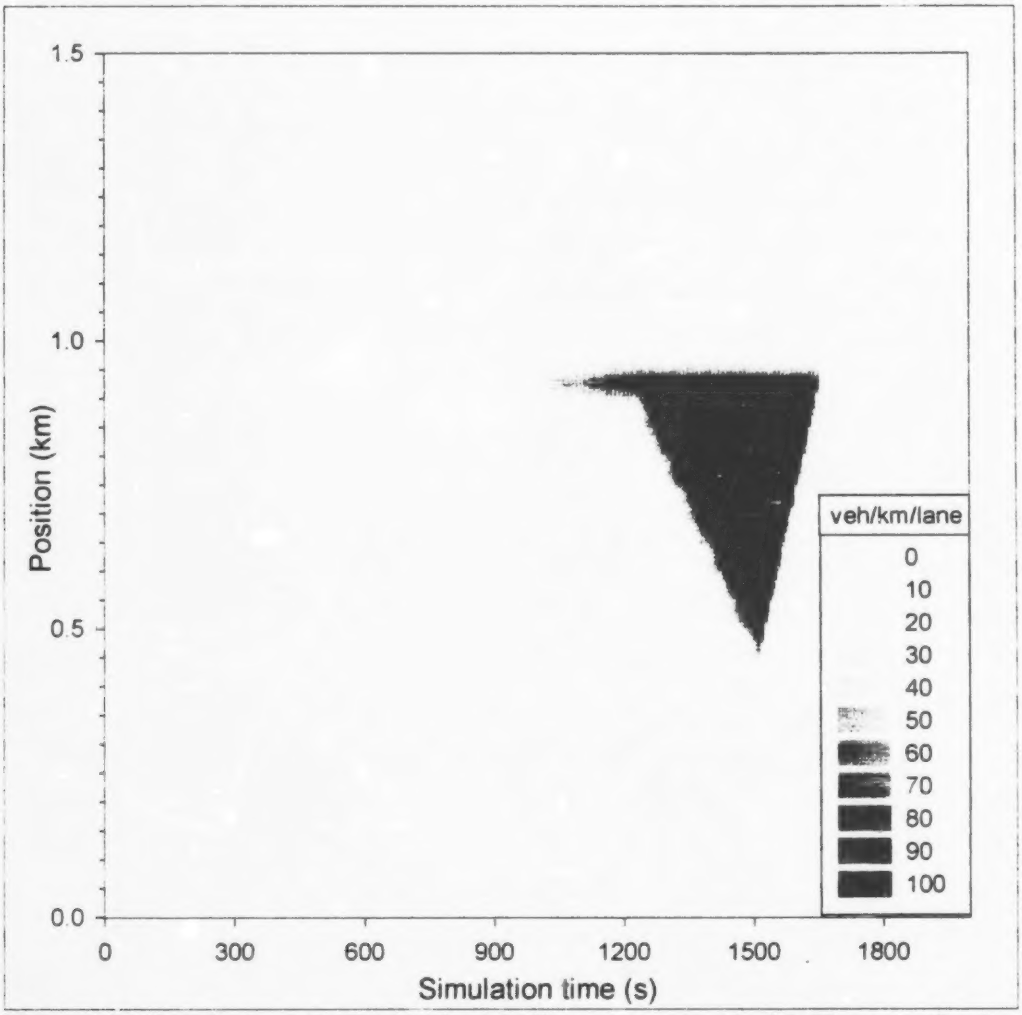


Figure 4.14 Density from CellNetLoad on the linear test network with 50-m arcs (2-s time step) and a reduced capacity at the lane-drop

The predicted path travel times for the CellNetLoad simulations with specified reduced capacity, shown in Figure 4.15 are very similar to the benchmark results in Figure 4.7, except for an underestimation of the path travel time during uncongested conditions.

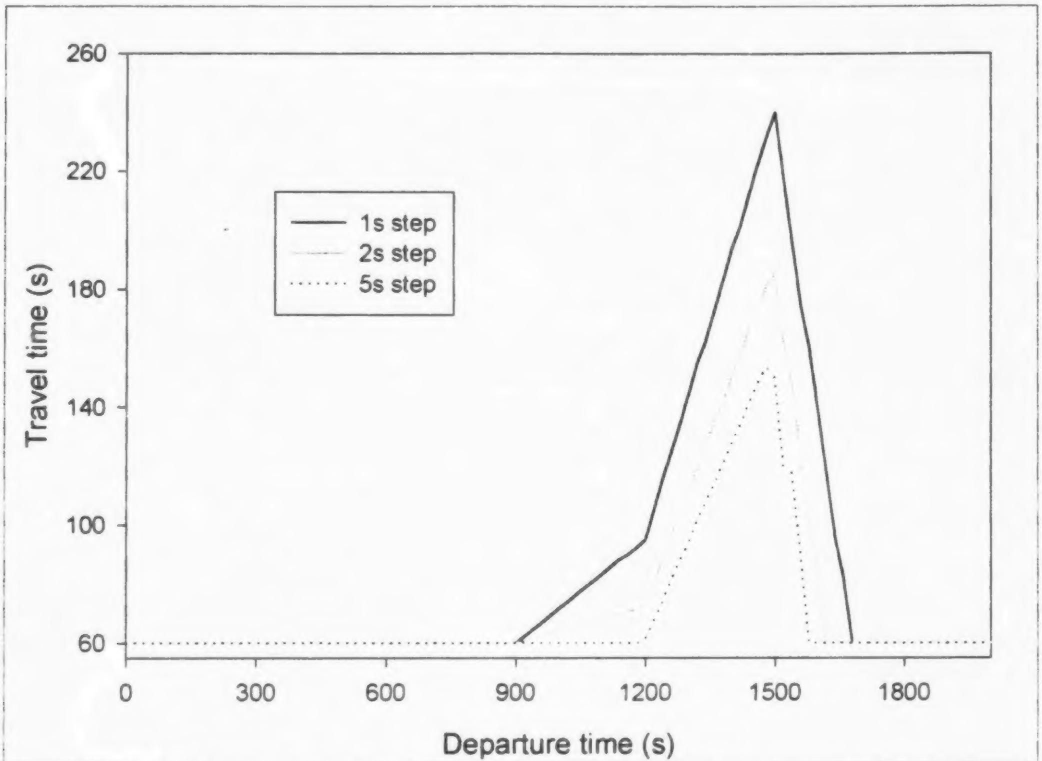
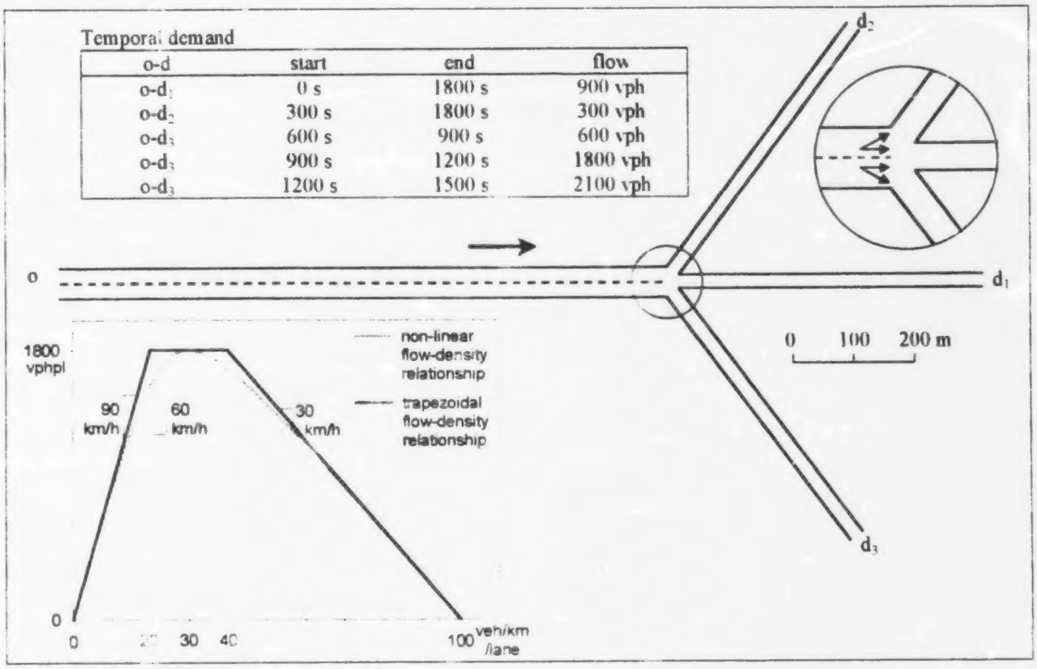


Figure 4.15 Path travel times from CellNetLoad on the linear test network with a reduced capacity (for time steps 1s and 2s)at the lane-drop

4.4 Diverge

The diverge network, shown in Figure 4.16, is identical to the linear network with the addition of two arcs to two additional destinations, so that a two-lane roadway splits at 1.0 km into three single-lane arcs each 0.5 km long. The diverge is configured such that through vehicles may use either lane, but turning movements may not cross from the far lane. The angle of the turn has no effect on the vehicle speeds in the simulation models considered here.



Temporal demand

o-d	start	end	flow
o-d_1	0 s	1800 s	900 vph
o-d_2	300 s	1800 s	300 vph
o-d_3	600 s	900 s	600 vph
o-d_3	900 s	1200 s	1800 vph
o-d_3	1200 s	1500 s	2100 vph

o
d_1
d_2
d_3
0 100 200 m
1800 vphpl
0
90 km/h
60 km/h
30 km/h
non-linear flow-density relationship
trapezoidal flow-density relationship
0
20
30
40
100
veh/km /lane

Figure 4.16 Diverge test network

The demand is increased each 300 s as in the linear network. A demand of 900 vph for destination 1 begins at the start of the simulation. A 300-vph demand is added for destinations 2 at 300 s and again for destination 3 at 600 s. This base of 1800 vph total demand continues until 1800 s. From 900 to 1200 s the right turn demand is increased by 1200 vph to 1800 vph, which is the saturation flow of the right lane. From 1200 to 1500 s the same right turn demand is increased by 1500 vph, so that the right lane is oversaturated at the diverge by 300 vph. The demand is returned to the base flow rate of 600 vph to destination 3 from 1500 to 1800 s. The demand then clears the network before the end of the simulation at 2000 s.

The diverge network is used to evaluate the diverge logic of the cell-transmission model with comparisons to a benchmark from INTEGRATION. The diverge logic distributes the discharge capacity of the diverge cell to each stream. This distribution is reflected in the flow into each of the three single-lane arcs, which was observed in 60-second intervals. The effect of the time-space discretisation on the diverge logic was also evaluated by using time steps of 5 s and 20 s, which correspond to cell lengths of 125-m arcs and 500-m arcs.

4.4.1 INTEGRATION simulation results

The flows for each stream at the diverge are displayed in Figure 4.17. The colour of the lines corresponds to a stream: black for the through movement, dark grey for the left-turn movement, and light grey for the right-turn movement. The thick lines represent demand at the origin. The thin, solid lines and dotted lines are the observed flow at the diverge for the network of 500-m arcs and 125-m arcs, respectively. Special attention is given to the analysis of the left turn movement, since it experiences a saturation demand and an oversaturation demand, as shown by the thick, light grey line.

The flows are spread evenly on the two lanes at the origin. Therefore, mandatory lane changes must occur for those right- and left-turning vehicles which enter the network on a lane from which they cannot turn. These crossing lane changes are known as weaving.

The impact of weaving was first at 720 s when some fluctuation in the through flow occurred, despite a volume to capacity ratio of 0.5. Once the 1800-vph right-turn demand reached the diverge at 960 s, there is a notable gap between the demand and the observed flow for all streams. This gap is a capacity reduction due to weaving. The queue begins in the right lane only. Left-turning vehicles continue to flow past the queue, and the through vehicles also overtake the queued right turners. Obviously, the overtaking violates first-in-first-out discipline. Eventually, the queue in the right lane spills over to the left lane, causing the left-turning and through streams to experience a larger capacity reduction.

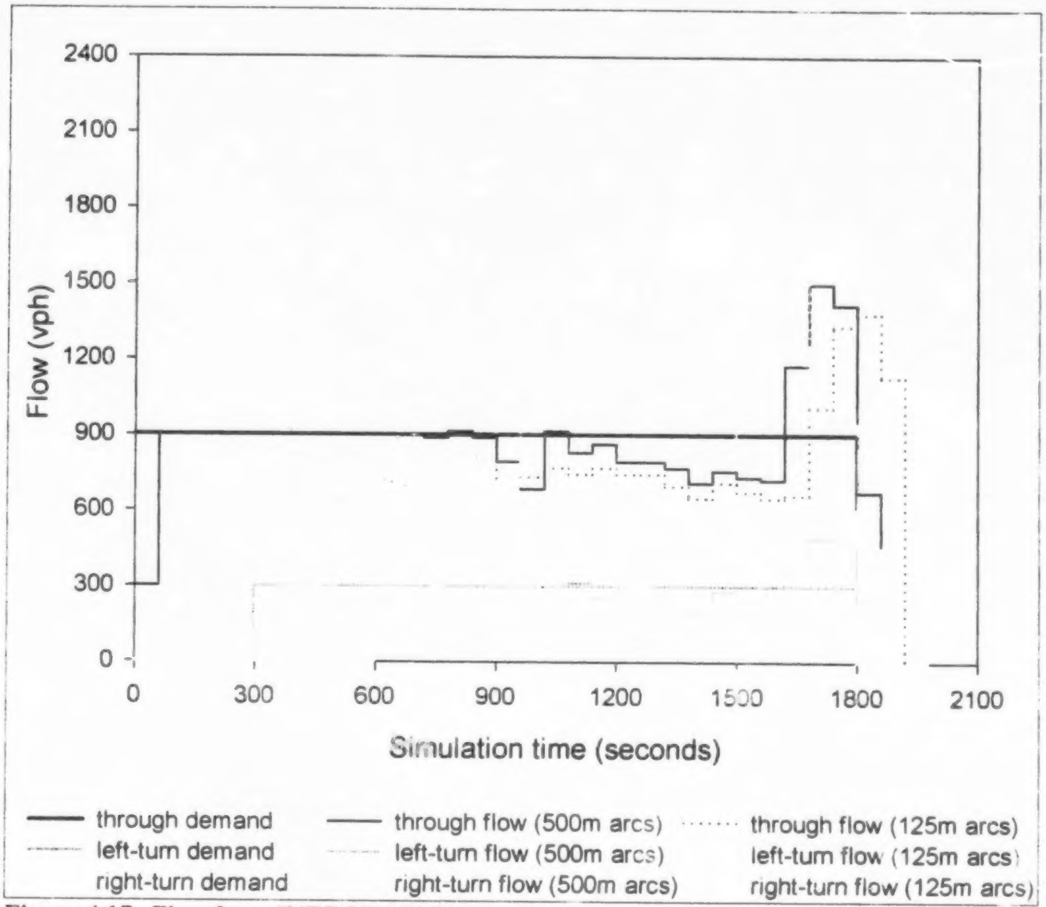


Figure 4.17 Flow from INTEGRATION at a diverge

The capacity reduction is more severe for the simulation with short arcs because vehicles must perform mandatory lane changes on the two arcs upstream of the diverge. Therefore, drivers reacted more aggressively when forced to seek gaps within 250 m, instead of the entire 1.0-km approach.

At 1500 s the base demand flows resumed, and the congestion began to clear. The total clearance flow was 3600 vph, the composition of which depended on the composition of the queue. The clearance took longer for the network with short arcs because the queue was larger.

4.4.2 CellNetLoad simulation results

The diverge logic of the cell-transmission model distributes the discharge capacity to the streams according to the occupancy of the diverge cell. Traditionally, the cell-transmission model enforces first-in-first-out discipline. This was accomplished in CellNetLoad by examining only the cohort at the front of the cohort queue in the diverge cell. During congestion, the first cohort may exceed the discharge capacity of the cell. In this case, the cohort is split using bucket rounding of the packets such that the first cohort exactly equals the discharge capacity. This implementation permits overtaking within a cohort. The first-in-first-out condition is enforced once some remaining packet in the first cohort cannot be discharged due to the local demand and local supply constraints.

Figure 4.18 shows the flows at the diverge observed in CellNetLoad. The flows for separate simulation runs are shown with different line styles. The thin, solid lines represent the observed flow for a 20-s time step (500-m cells) and the dotted lines correspond to a 5-s time step (125-m cells).

The saturated right-turn demand was permitted to flow through the diverge without causing any congestion, because the cell-transmission model applies no capacity reduction due to weaving. However, a queue began when the oversaturated right-turn demand reached the diverge. The first cohort of the oversaturated demand to arrive at the diverge contained 18.33 vehicles: 1.67 left-turning vehicles, 5.00 through vehicles, and 11.67 right-turning vehicles. The local demand was constrained by the discharge capacity of 20 vehicles, which was distributed as 1.82 vehicles to the left-turn movement, 5.45 vehicles to the through movement, and 12.73 vehicles to the right-turn movement. The local supply for each stream was always 10 vehicles, since the single-lane arcs remained uncongested throughout the simulation. Therefore, the left-turners and through vehicles were stored in the cohort list ready for discharge, but 1.67 right-turning vehicles of the original 11.67 vehicles remained in the first cohort.

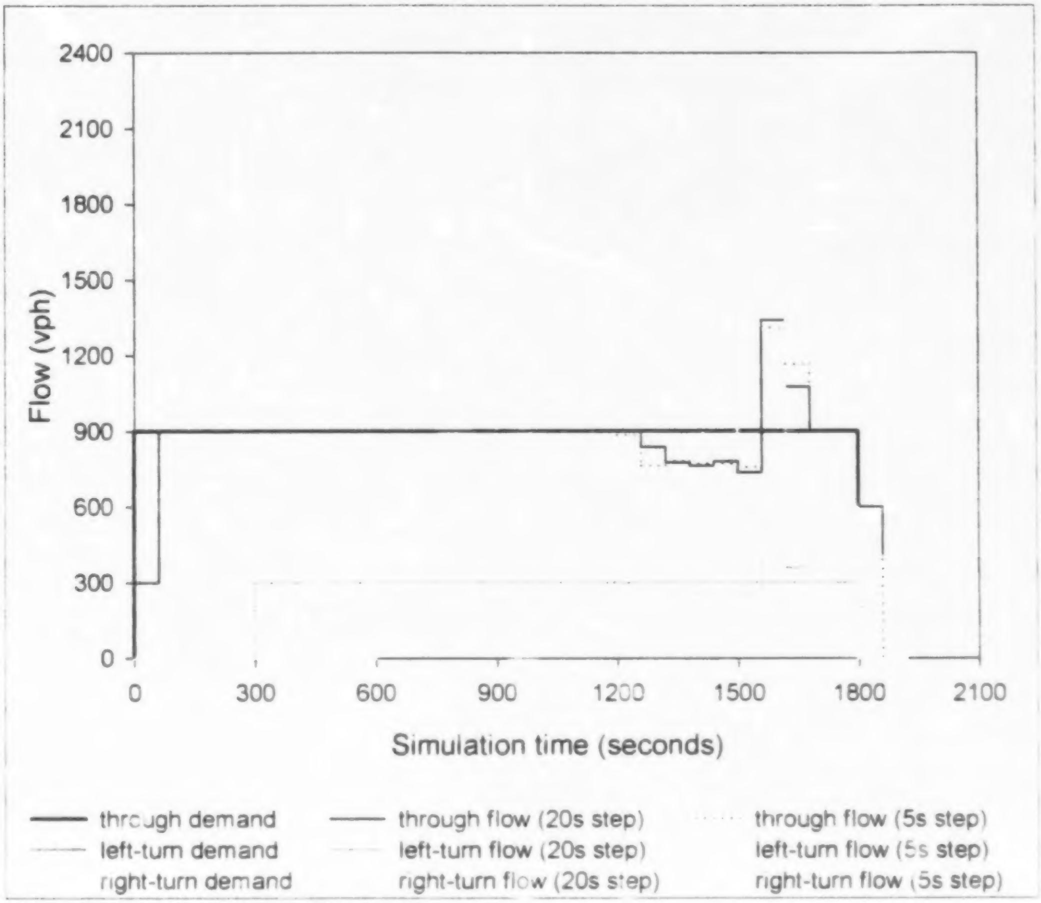


Figure 4.18 Flow from CellNetLoad at a diverge

On the next time step, the first cohort of 1.67 right-turning vehicles were stored in the cohort list ready to be discharged. The remaining local demand of 18.33 vehicles was exactly equal to the next cohort, which became the first cohort. The remaining discharge capacity was distributed such that no constraints were encountered, but the remaining local supply was only 8.33 vehicles for the right-turn movement. So, 3.33 right-turning vehicles were left in the first cohort.

The following time step discharged the first cohort of 3.33 right turners, then considered the next cohort. This cohort exceeded the remaining local demand of 16.67 vehicles, so it was split. Now, the first cohort contained 1.58 left turners, 4.55 through vehicles, and 10.61 right turners. The second cohort contained the other 0.09 left turners, 0.45 through vehicles, and 1.[illegible]6 right turners.

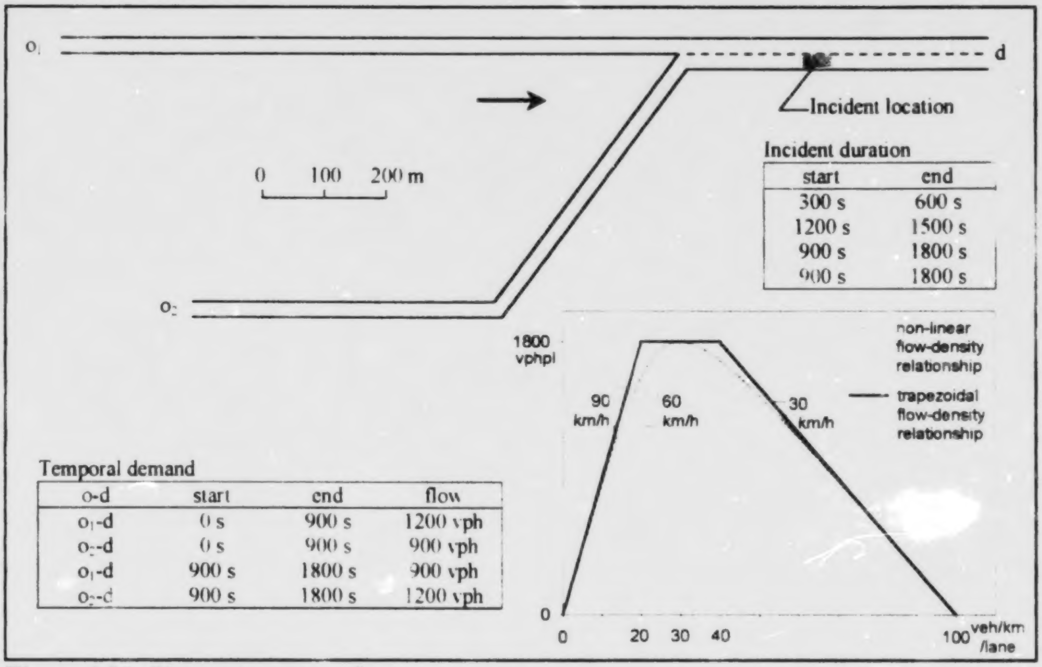
The right turners in the first cohort could not fully discharge because the remaining local supply was only 6.67 vehicles. The second cohort was blocked by the first cohort. Note that the second cohort contained all movements.

Some of the demand from each movement was accumulated in the queue. This explains the gaps shown in Figure 4.18 between the demand and the observed flow for the 20-s time step. The gaps appear earlier for the simulation with a 5-s time step because it takes the same number of time steps for the right turners to block the other streams. However, note that the degree of blockage reaches the same level for both time steps. This level of blockage is not stable because it depends on the size of the front cohort. After the oversaturation demand ends, the queue clears according to the proportion of each stream in the cohort at the front of the queue.

The results from the CellNetLoad differ from INTEGRATION in two significant ways. Firstly, the capacity drop due to weaving is not captured by CellNetLoad. This could be corrected by specifying a reduced capacity in the input, but it is difficult to predict. Secondly, overtaking is limited in CellNetLoad. The first-in-first-out discipline may be appropriate for single lane approaches, but it does not permit overtaking on multi-lane approaches.

4.5 Merge

The merge of two 1.0-km, single-lane arterials into a two-lane arterial is shown in 0. Though the arcs are configured to appear like a highway on-ramp, the angles have no impact on the simulations. Nevertheless, for the sake of simplicity the upper arc is referred to as the mainline, and the lower arc is called the on-ramp.


o$_1$
d
o$_2$
Incident location
0
100
200 m
1800 vphpl
90 km/h
60 km/h
30 km/h
non-linear flow-density relationship
trapezoidal flow-density relationship
0
0
20
30
40
100
veh/km /lane

Incident duration

start	end
300 s	600 s
1200 s	1500 s
900 s	1800 s
900 s	1800 s

Temporal demand

o-d	start	end	flow
o_1-d	0 s	900 s	1200 vph
o_2-d	0 s	900 s	900 vph
o_1-d	900 s	1800 s	900 vph
o_2-d	900 s	1800 s	1200 vph

Figure 4.19 Merge test network

The merge logic determines the distribution of the local supply of the downstream arc to the upstream arcs. The merge logic has an impact on the traffic flow only if the downstream local supply constrains the upstream flows. This was accomplished by specifying the occurrence of an incident on the downstream arc that blocks flow in one lane. A lane reduction at a merge is common, but the use of an incident permits a temporary lane drop.

Lane drops usually occur on one side of the roadway, leading to an asymmetry. This asymmetry was investigated by specifying asymmetric demand flows from the origins, then switching the demands from the origins. The mainline demand was 1200 vph, and the on-ramp demand was 900 vph during the first 900 s of the simulation. These demand flows were switched during the subsequent 900 s.

One incident was specified during the middle 300 s of each set of demands. So, the first 300 s were uncongested flow. Then congestion occurred during the next 300 s, which cleared during the following 300 s. This cycle repeated itself for the second 900 s of the simulation.

The flows of both streams at the merge were observed in 60-s collection intervals and presented in graphs. The effect of arc length on the lane -changing logic of INTEGRATION, and the effect of the time-step on CellNetLoad were examined.

4.5.1 INTEGRATION simulation results

In INTEGRATION the incident occupied 50 m of the right lane 200 m downstream from the merge. So, the on-ramp flow was forced to switch lanes between the merge and hard wall, which is encountered some distance upstream from the incident. The observed flows are a direct result of the lane-changing. Because lane-changing is a gap-acceptance process, it was found that the distribution of the gap size, which is the headway, played a crucial role. Thus, the simulations were run using a constant departure headway, then repeated specifying exponentially distributed departure headways. The results are displayed in 0 and Figure 4.21. The black and grey lines correspond to the mainline and on-ramp, respectively. The thick, solid line is the demand. The thin, solid line is the flow observed for the network consisting of 500-m arcs. The dotted line is the flow observed for the network consisting of 125-m arcs.

The flows lag behind the demand due to the travel time to the merge, but eventually they are equal to the demand flow until they reach the incident. The aggressiveness of the lane-changes depends on the distance from the hard wall. The hard wall is located on the arc which contains the incident. So, the hard wall was located 125 m from the merge for the network with short arcs, but it was closer to the merge for the network with longer arcs. The lane-changes were more aggressive for the simulation with 500-m arcs because the window of opportunity for the lane change was smaller.

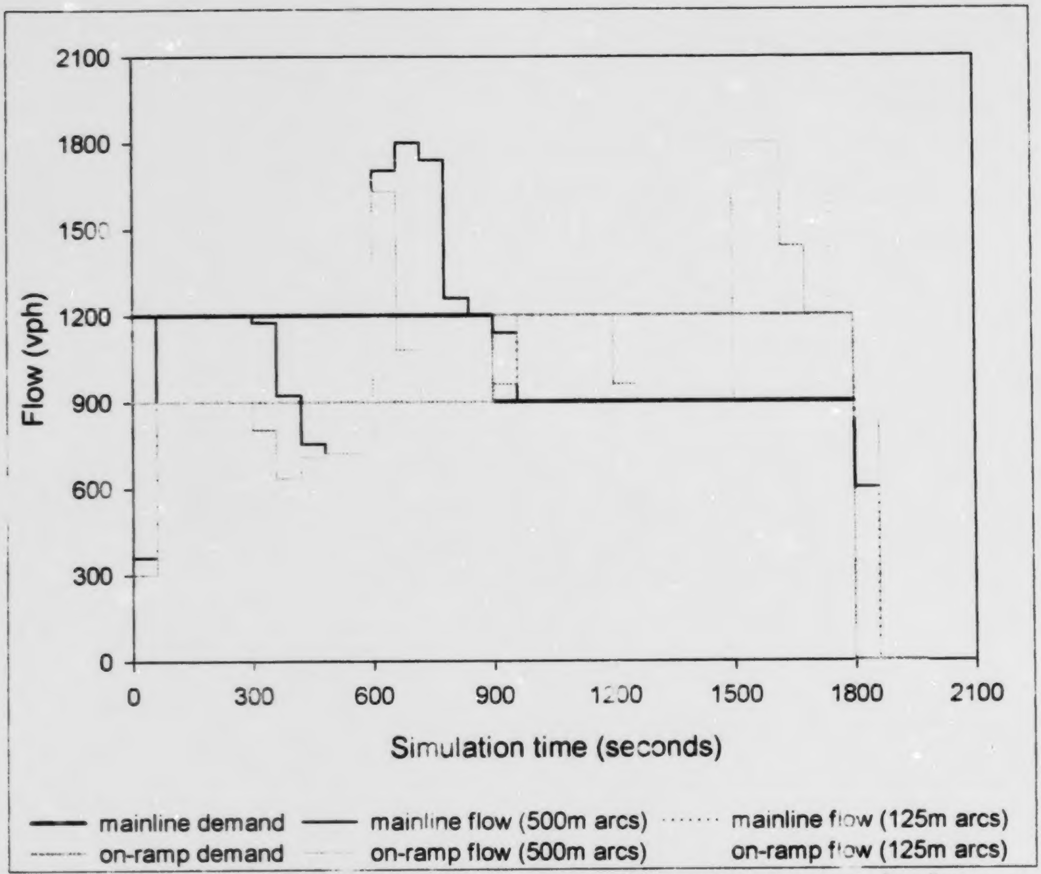


Figure 4.20 Flow from INTEGRATION at a merge with constant departure headways

For the simulation with 500-m arcs, the aggressive lane changing of the on-ramp flow forces the mainline flow into a congested state. With both lanes congested, the departures from the front of the queue alternate from each lane. One may expect the departures to occur every 2 seconds, since the saturation flow rate is 1800 vphpl. However, the vehicles must accelerate, which causes a reduction in the capacity to 1440 vphpl. This reduced capacity is shared by both lanes once an equilibrium is reached. Both flows converge to 720 vph, as shown on Figure 4.20. Queues grow on both lanes, spilling back to both upstream arcs.

The simulation with 125-m arcs has less aggressive lane changing. The mainline flow is unaffected, remaining equal to the demand at 1200 vph with 3.0-s headways. (The dotted line is covered by the thick, solid line in Figure 4.20.) The on-line flow accepts every second gap, leading to a departure from the queue every 6.0 s, and an observed flow of 600 vph. A queue grows in the right lane, spilling back to the on-ramp.

The queues clear after the incident ends at 600 s. The demands are switched at 900 s. A second incident occurs in the right lane at the same location at 1200 s.

The on-ramp flow accepts each gap in the mainline flow for the simulation with 500-m arcs. The mainline flow is unaffected, remaining equal to the demand at 900 vph. The on-ramp flow is reduced to 900 vph, causing a queue to grow at 300 vph in the right lane. The on-ramp flow is less aggressive for the simulation with 125-m arcs, accepting three gaps and declining the fourth, leading to an observed flow equal to 675 vph, which is three-quarters of 900 vph.

This deterministic gap acceptance is unrealistic because the gaps are of constant size. A more realistic distribution of the available gaps in the mainline flow is achieved by specifying exponentially distributed departure headways. This randomness means that the mean number of departures remains equal to the specified flow rate, but the number of departures during any observation interval may vary as a Poisson process. In order to reduce the stochasticity in the observations, the simulation was run ten times with different seeds. The observations were averaged, and displayed in Figure 4.21.

Even with results averaged from ten simulation runs, the observed flow is quite jumpy. The first incident shows a similar pattern as in Figure 4.20. In the second incident, however, the mainline flow is somewhat congested due to the occasional acceptance of a small gap. For the simulation with 500-m arcs, the on-ramp flow drops to an observed flow that is the same as in the first incident. For the simulation with 125-m arcs, the mainline flow is less congested, because the on-ramp flow is less aggressive.

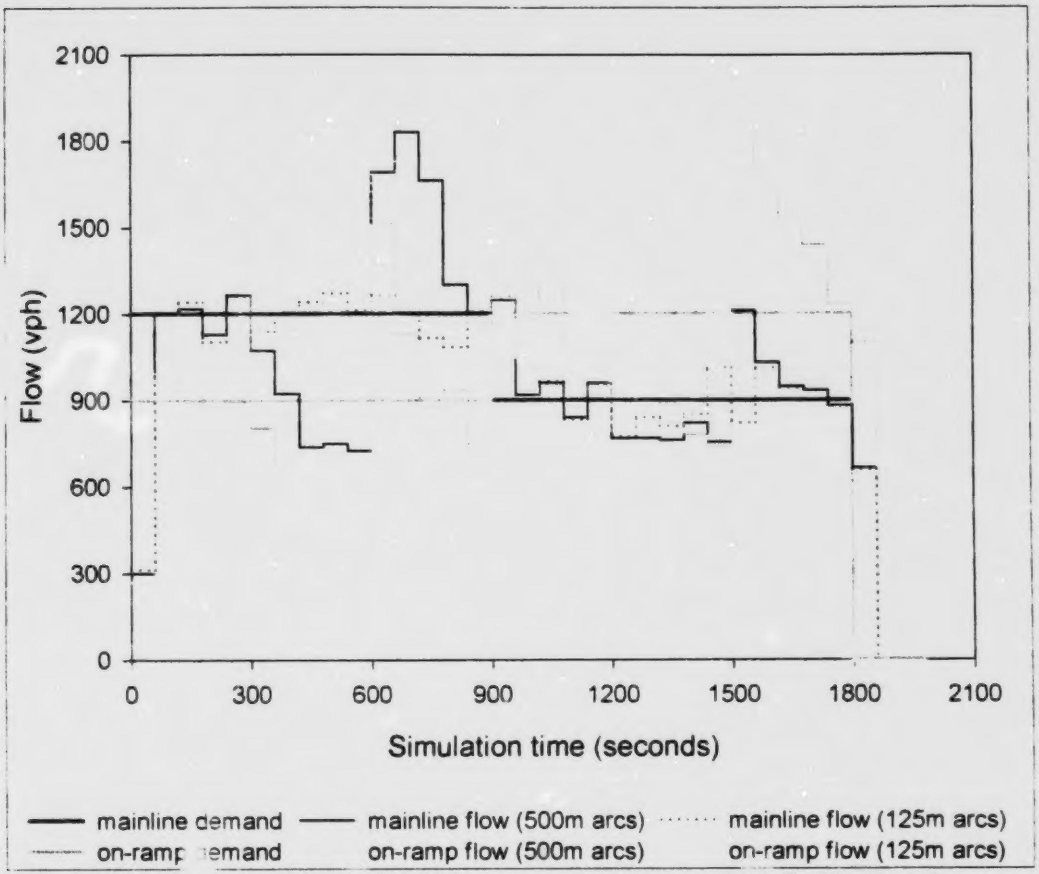


Figure 4.21 Flow from INTEGRATION at a merge with exponentially distributed departure headways

The conclusion, which can be drawn from these simulation results, is that INTEGRATION explicitly models the asymmetry due to the lane-reduction in the right lane of the two-lane arc. Also, the flows at this particular merge were completely dependant on the aggressiveness of the on-ramp vehicles in changing lanes.

4.5.2 CellNetLoad simulation results

The reduction in flow capacity and storage capacity due to an incident is applied to each cell which happens to fall within the specified location of the incident. The flow capacity is reduced by the severity of the incident. The storage capacity is also reduced according to the severity of the incident, but only on the portion of the cell length that falls within the incident location. The axes of the equilibrium flow-density relationship are scaled using these reductions to capacity flow and jam density. This representation of an incident implies a symmetric treatment across the width of a an arc. The asymmetry of the lane drop is not reflected in the simulation results.

The symmetry of the observed flows is evident from an examination of Figure 4.22. The congestion began with the first incident at 300 s. For the simulation with a 20-s time step, the incident occurred on the first 500-m long cell after the merge, so the local supply was immediately restricted to 10 vehicles (one lane of 1800 vph for 20 s) at the merge cell. The 5-s time step reduced the cell length to 125 m. The incident occurred on the fourth cell after the merge. The queue had to spill back the length of three cells and into the fourth before the supply was restricted, which took almost three minutes in the simulation.

The local supply was shared at the merge in proportion to the upstream demand. At first, the front cohort of the mainline was larger than that of the on-ramp, so it had a larger inflow to the merge cell. The queue grew faster on the freeway, exceeding the local demand by 380 s for the simulation with a 20-s time step. At 380 s the mainline front cohort is split into 6.188 vehicles (the local demand) and 0.479 vehicles (the rest). When the small split cohort reaches the front of the cohort queue on the mainline during the second iteration at 400 s, the partial demand is very small. This temporarily favours the on-ramp flow, which prevents the on-ramp queue from ever exceeding the local demand. The total flow through the merge is always equal to 1800 vph, but the sharing of this capacity is unstable, as a result of the splitting of cohorts.

The reduction to the flows is mirrored during the second incident, since the incident is implicitly modelled as a uniform, hence symmetric, blockage across all lanes, while the demands were simply switched for the mainline and on-ramp.

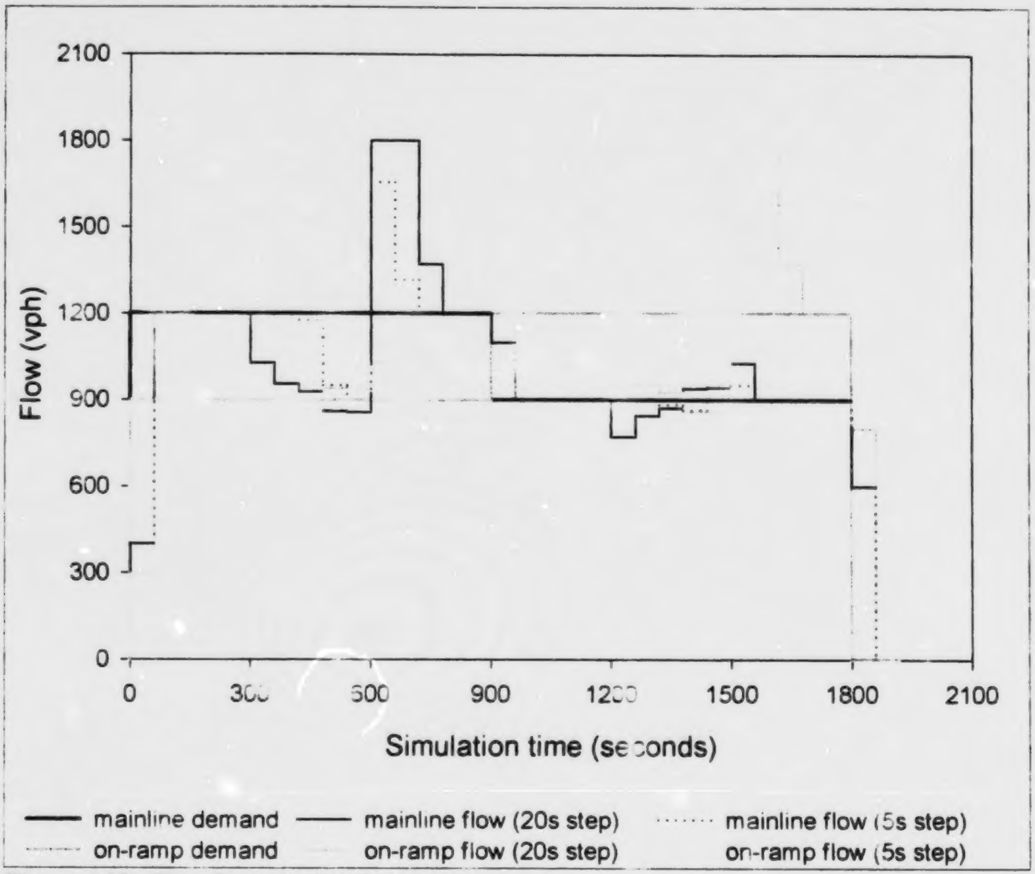


Figure 4.22 Flow from CellNetLoad at a merge

The flow reductions occurred later in the simulation with the 5-s time step, because the queue took nearly three minutes to reach the merge cell. No instability was observed for the 1-minute aggregation of inflow. The convergence to 900 vph took less than 2 minutes.

4.6 Intersection

A hypothetical signalised intersection of a two-lane arterial and a single-lane minor street, shown in Figure 4.23, was used to evaluate the cell-transmission model. The equilibrium flow-density relationships of the major and minor streets, also shown in Figure 4.23, were revised to be more representative of an urban arterial and a minor street. The major street approaches were 400 m, and the minor street approaches were 267 m, so that all approaches had a 20-s travel time at the free speed. The signal timing plan in the lower right of Figure 4.23 was a simple two-phase plan with 4-s yellow intervals. Because no right-turn-on-red was permitted, the major and minor roads operated independently of one another.

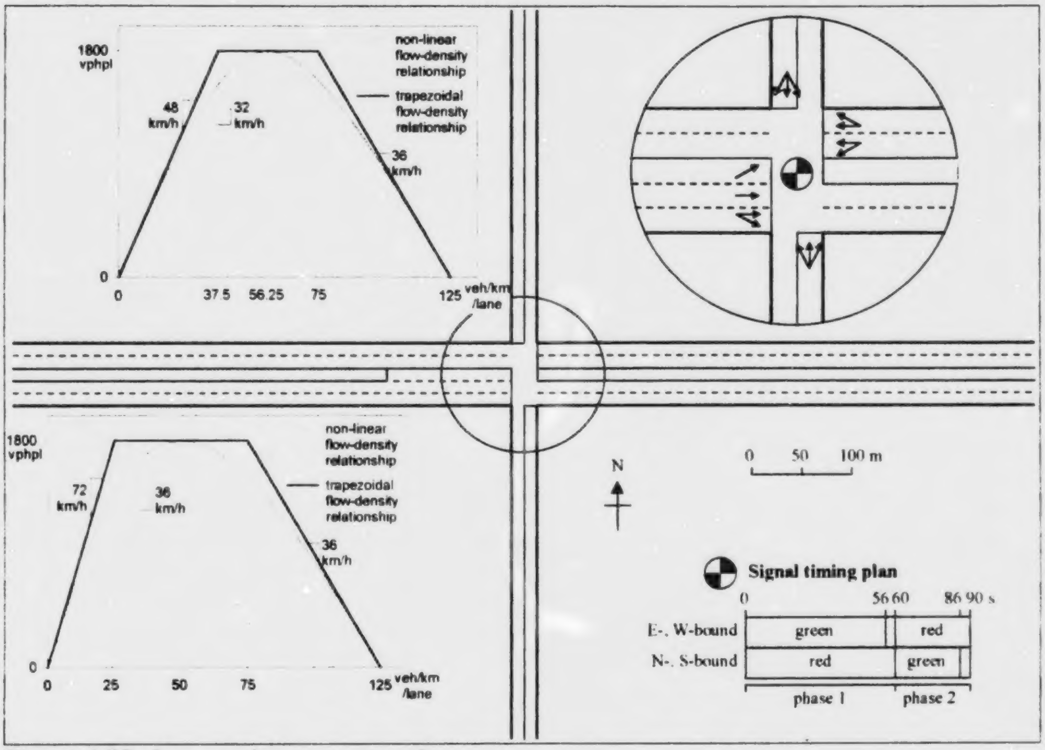

1800 vphpl
non-linear flow-density relationship
trapezoidal flow-density relationship
48 km/h
32 km/h
36 km/h
0
0
37.5
56.25
75
125
veh/km /lane
1800 vphpl
non-linear flow-density relationship
trapezoidal flow-density relationship
72 km/h
36 km/h
36 km/h
0
0
25
50
75
125
veh/km /lane
N
0
50
100 m
Signal timing plan
0
56
60
86
90 s
E-, W-bound
green
red
N-, S-bound
red
green
phase 1
phase 2

Figure 4.23 Intersection test network

The demands were constructed with permissive left turns on all approaches except the northbound approach. The lane-striping was designed to model a variety of situations for the left-turners, which were always opposed permissive movements. Eastbound left-turners used a left-turn bay, separated from the through movement. The westbound left-turners shared a lane with the through movement, which also shared the right lane with the right-turn movement. The southbound left-turners shared a single lane with right-turners.

Three scenarios were modelled. The demands for the heavy flow scenario were carefully selected using the animation of INTEGRATION to create mild oversaturation. The demands were reduced to half for the scenario with light flow. The third scenario used the same demand as the heavy flow scenario, but included a 50-m incident on the right lane of the eastbound arterial 250-m downstream of the intersection, which was intended to create greater oversaturation. The demands in each scenario remained the same for all simulation tests, and are included in the figures which contain the simulation results.

The simulation results are displayed on a rough map of the intersection. Observations were collected during each cycle of the signal, then averaged for the second through twenty-first cycles. The first phase was ignored because the vehicles had initially not reached the intersection. The three diagrams are each labelled with

- the demand for each movement in vehicles per hour;
- the average observed flow of each movement in vehicles per hour;
- the average observed density of each approach in vehicles per kilometre per lane; and
- average observed travel time of each approach in seconds of the vehicles which entered the approach during the cycle.

4.6.1 INTEGRATION simulation results

INTEGRATION models traffic signals explicitly. A small start loss occurs at the beginning of the green phase due to the acceleration of the first few vehicles in the queue, but does not include a reaction time. There is no end gain, since vehicles are not permitted to pass through a yellow indication. One notable exception is the left turn sneaker, a vehicle which departs on the yellow indication to simulate the clearance of a left turner who had already crossed the stop line during the green indication. Multiple left-turn sneakers may use a single yellow indication if it is long enough. All other permissive left turns must pass through gaps in the opposing traffic stream(s).

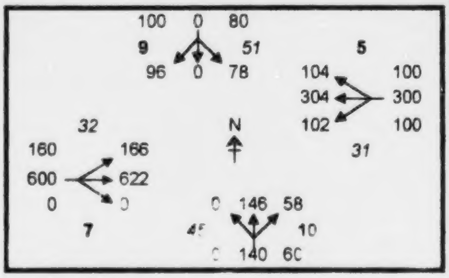
The identification of conflicting movements, priority, and the gap acceptance logic is more thoroughly described in Velan and Van Aerde (1997).

The INTEGRATION simulation results are shown in Figure 4.24 for all three scenarios. The arrows represent movements. Each arrow is labelled with the demand at the beginning and with the average observed flow at the end. The average observed density, shown in bold typeface, is the total density of all movements on the approach. The average observed travel time, shown in italics, is a weighted average of all movements on the approach. The above values are also included in tables to the right of the map, with corresponding typeface.

The average observed flows do not match the demand exactly. In the scenario with light demand flow, this deviation of less than 5 percent is explained by the exponentially distributed departure headways. Deviations larger than 5 percent are more likely due to oversaturated movements which accumulated a queue during the simulation. In the scenario with heavy flow, the southbound approach was oversaturated, as were the left-turn movements on the eastbound and westbound approaches. The incident caused even more oversaturation. Surprisingly, the westbound approach was greatly oversaturated, while the flow on the eastbound approach actually increased. The northbound and southbound approach flows were unaffected by the incident.

The density is an indication of the queuing. The doubling of the demand more than doubled the average density. The density increased by a factor of 3.9, 3.8, 7.4, and 2.4, which seems to be correlated to the proportion of the approach volume that is opposed left turns: 0.21, 0.20, 0.44, and 0.00. (Unless otherwise stated, all values will be given in the same order as the tables: eastbound, westbound, southbound, and northbound.)

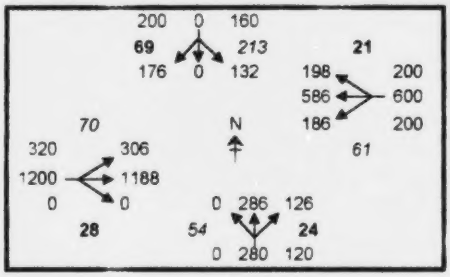
Light Flow


100
0
80
9
51
96
0
78
5
104
304
102
100
300
100
32
160
600
0
166
622
0
N
31
7
0
146
58
45
10
0
140
60

approach	demand (vph)		
	left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	166	622	0	**7.2**	*32*
W bound	102	304	104	**5.4**	*31*
S bound	78	0	96	**9.4**	*51*
N bound	0	146	58	**9.9**	*45*

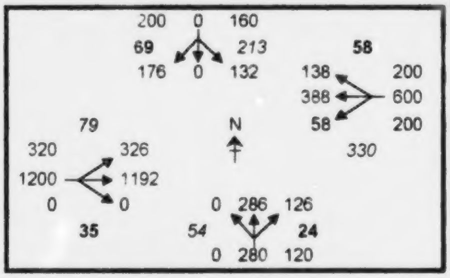
Heavy flow


200
0
160
69
213
176
0
132
21
198
586
186
200
600
200
70
320
1200
0
306
1188
0
N
61
28
0
286
126
54
24
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	306	1188	0	**28.2**	*70*
W bound	186	586	198	**20.7**	*61*
S bound	132	0	176	**69.3**	*213*
N bound	0	286	126	**23.9**	*54*

Heavy flow with an incident


200
0
160
69
213
176
0
132
58
138
388
58
200
600
200
79
320
1200
0
326
1192
0
N
330
35
0
286
126
54
24
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	326	1192	0	**34.7**	*79*
W bound	58	388	138	**58.1**	*330*
S bound	132	0	176	**69.3**	*213*
N bound	0	286	126	**23.9**	*54*

Figure 4.24 Simulation results from INTEGRATION of an intersection

The interactions of the movements is complicated, even for this simple intersection. The queuing on the southbound approach was due to the plug created by the left turn vehicles when opposed by a heavy northbound flow. The incident increases the density on the eastbound approach from 28.2 to 34.7 veh/km/lane, though the observed flow actually increased. The eastbound platoon was forced to squeeze past the incident in a single lane, which caused a shockwave to travel from the incident back to the stop line. The flow across the stop line fell below the saturation flow rate. Although the queue was slower to discharge, the green indication was long enough to permit clearance of the queue. Because the eastbound queue discharged during nearly the entire phase, few acceptable gaps were available to the westbound left turn movement. The westbound left turns consisted entirely of sneakers. The queue of westbound left turners grew to 300 m and spilled over to block the through and right-turn movements, then spilled back to the origin. The westbound density was actually 78 veh/km/lane, but only on the left lane and the upstream quarter of the right lane, yielding a reported average density of 58.1 veh/km/lane. This blockage of the westbound approach permitted the eastbound left turns to proceed unopposed, which explains the increase in the observed flow.

The free-speed travel time was 20 s for each approach. An analytical calculation of the queuing delay under light demand flow accounts for 7.9, 7.9, 26.1, and 26.3 s/veh. The remaining travel time of approximately 4, 3, 5, and –1 seconds can be attributed to lane-changing, left-turn delay, and random delay, which is the temporary increased or decreased congestion—and consequent positive or negative change in travel time—due to exponentially distributed departure headways.

For the heavy demand flow, accounting for the free-speed travel time and average queue delay leaves a remaining travel time of 40 s/veh on the eastbound approach, which is attributed to random delay, left turn delay, and spillover from the left-turn bay which impeded the through movement. The same factors accounted for 32 s/veh on the westbound approach. The left turn plugging of the southbound approach accounted for 164 s/veh. The northbound approach had only 4 s/veh of random delay.

The incident caused an additional 9 s/veh of delay to the eastbound approach, which is explained by the slower discharge of the queue due to the shockwave, but less left-turn delay. The westbound approach was oversaturated, causing a 301 s/veh increase in the average travel time. The northbound and southbound approaches were unaffected by the incident.

4.6.2 CellNetLoad simulation results

The traditional cell-transmission model, as described in chapter 2, defines cells as either ordinary, merge, or diverge. For the exiting flow to be dependent only on the density of that cell and the downstream cell(s)—a necessary property for a fully parallel implementation—several restrictive conditions must be satisfied. A diverge cell sends flow to ordinary cells only. A merge cell receives flow from ordinary cells only. This implies that an ordinary cell must be inserted for each stream at an intersection to receive the divergent flow from the approach, and to send the merging flow to the downstream arc. This ordinary cell may be placed on the approach. Such an intersection, shown in Figure 4.25, appears and behaves like a freeway interchange.

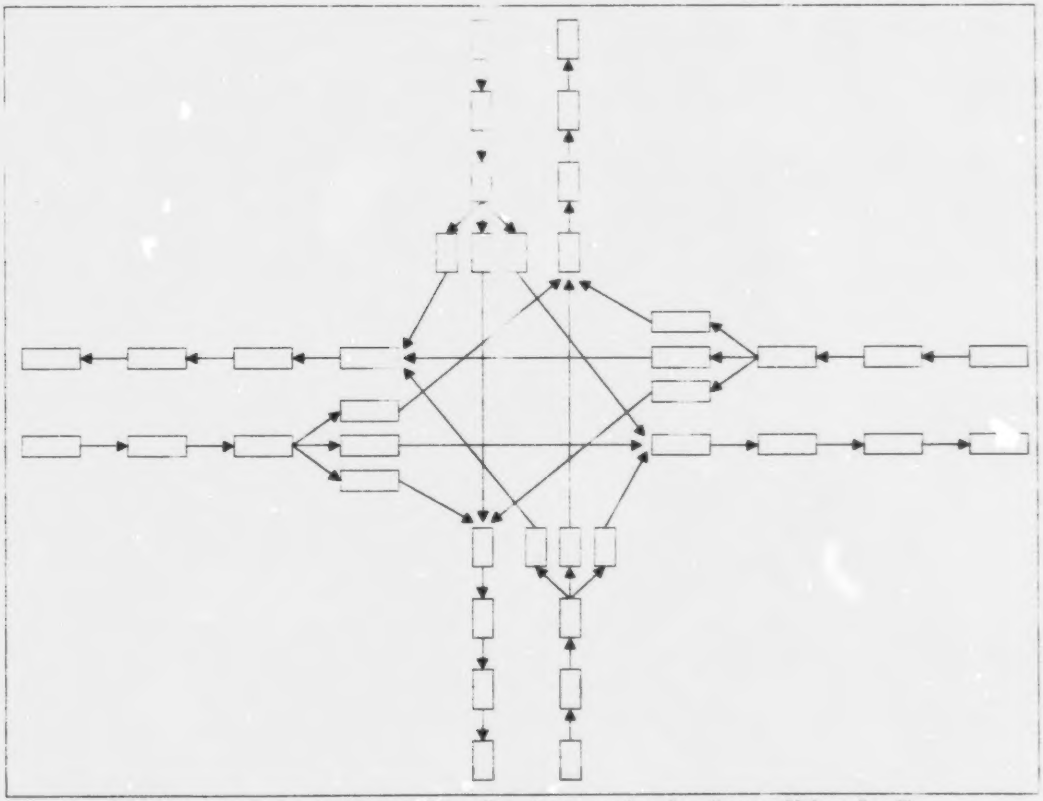


Figure 4.25 Interchange representation of an intersection by the traditional cell-transmission model

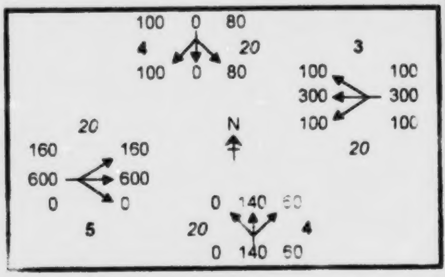
The capacity of the approach must be divided among the ordinary cells using a model such as the Highway Capacity Manual (HCM, 2000), as suggested by Ziliaskopoulos and Lee (1997). The distribution of the approach capacity among the streams depends on many factors, for example, capacities of opposed permissive movements, spillover of queues, and lane-changing.

Figure 4.26 shows the CellNetLoad simulation results for the intersection with an interchange-representation. In order to use a time step of up to 5 s, the diverge was placed 100 m upstream of the intersection on the arterial approaches, and 67 m upstream from the intersection on the minor street approaches.

The flow capacity of the ordinary cells was reduced according to the observed flows during the green indication in INTEGRATION. To limit spillback the storage capacity of the ordinary cells was generously made equal to the product of the jam density, the cell length, and the number of lanes with permission to discharge that stream. Shared lanes effectively doubled the storage capacity at the end of the westbound approach and tripled it at the end of the northbound and southbound approaches.

Signals were modelled statically, meaning that the capacity of each approach was reduced by the ratio of the green interval to the cycle length. This signal model permitted the simultaneous merging of streams from the arterial and minor street. Because the local supply on the downstream arcs was never a flow constraint, this undesirable feature is not reflected in the simulation results. The travel time contains only the free-speed travel time and any delay associated with oversaturated movements. Delay associated with lane-changing and queuing during the red phase is not modelled.

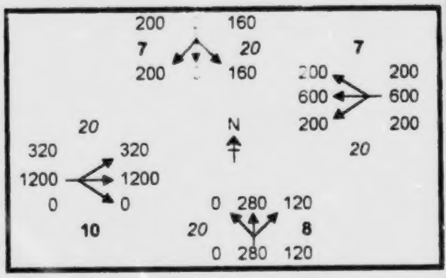
Light Flow


100
0
80
4
20
3
100
0
80
100
100
300
300
100
100
20
N
20
160
160
600
600
0
0
0
140
60
5
20
4
0
140
60

approach	demand (vph)		
	left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	160	600	0	4.8	20
W bound	100	300	100	3.5	20
S bound	80	0	100	3.7	20
N bound	0	140	60	4.2	20

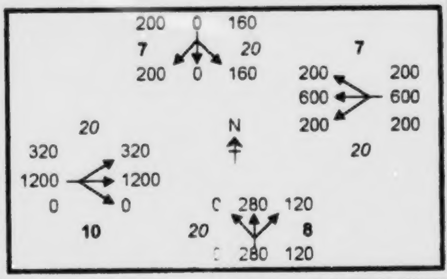
Heavy flow


200
160
7
20
7
200
160
200
200
600
600
200
200
20
N
20
320
320
1200
1200
0
0
0
280
120
10
20
8
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	9.7	20
W bound	200	600	200	6.9	20
S bound	160	0	200	7.5	20
N bound	0	280	120	8.3	20

Heavy flow with an incident

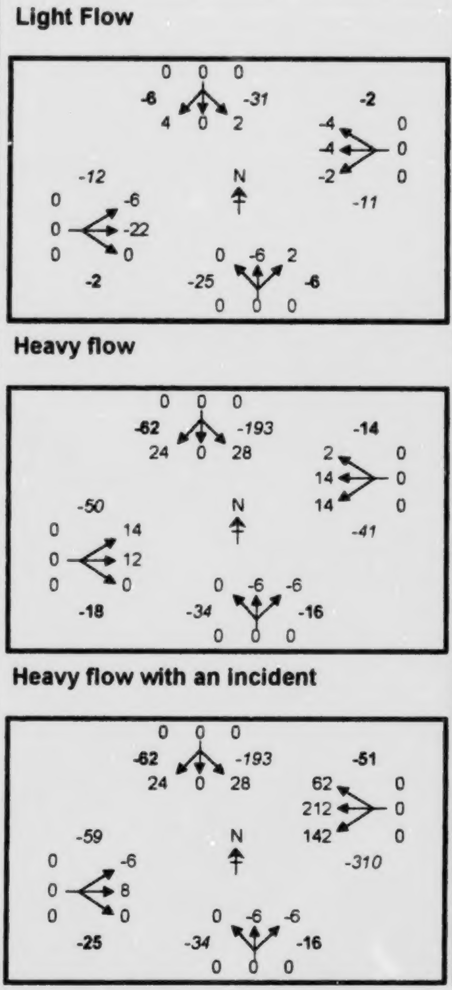

200
0
160
7
20
7
200
0
160
200
200
600
600
200
200
20
N
20
320
320
1200
1200
0
0
0
280
120
10
20
8
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	9.7	20
W bound	200	600	200	6.9	20
S bound	160	0	200	7.5	20
N bound	0	280	120	8.3	20

Figure 4.26 Simulation results from CellNetLoad of an intersection with interchange representation and static signals

A direct comparison is made in Figure 4.27 by subtracting the INTEGRATION values from those observed in CellNetLoad. The observed flows match very closely for the light flow and heavy flow scenarios, but the capacity reductions in the incident scenario were severe enough to cause spillback to the diverge. Thus, a double penalty was applied due to the incident: an externally applied capacity reduction, and an internal reduction due to the spillback at the diverge. This spillback could be eliminated by specifying an infinite storage capacity for the cells downstream from the diverge cell, but this would also eliminate all shockwaves from propagating across intersections. A realistic distribution of the storage capacity of shared lanes would only worsen the impact spillback to the diverge cell.


0 0 0
-6
-31
4 0 2
-2
-4 0
-4 0
-2 0
-12
N
-11
0 -6
0 -22
0 0
-2
0 -6 2
-25
-6
0 0 0
0 0 0
-62
-193
24 0 28
-14
2 0
14 0
14 0
-50
N
-41
0 14
0 12
0 0
-18
0 -6 -6
-34
-16
0 0 0
0 0 0
-62
-193
24 0 28
-51
62 0
212 0
142 0
-59
N
-310
0 -6
0 8
0 0
-25
0 -6 -6
-34
-16
0 0 0

Light Flow

approach	demand (vph) left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph) left	through	right	density (veh/km/ln)	travel time (seconds)
E bound	-6	-22	0	**-2.4**	*-12*
W bound	-2	-4	-4	**-2.0**	*-11*
S bound	2	0	4	**-5.7**	*-31*
N bound	0	-6	2	**-5.7**	*-25*

Heavy flow

approach	demand (vph) left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph) left	through	right	density (veh/km/ln)	travel time (seconds)
E bound	14	12	0	**-18.5**	*-50*
W bound	14	14	2	**-13.8**	*-41*
S bound	28	0	24	**-61.8**	*-193*
N bound	0	-6	-6	**-15.6**	*-34*

Heavy flow with an incident

approach	demand (vph) left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

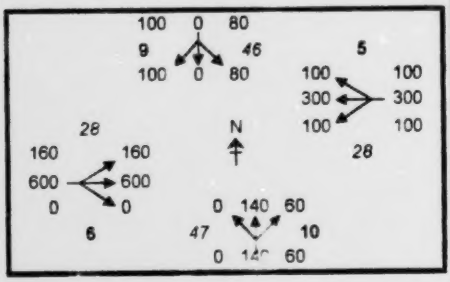
approach	volume (vph) left	through	right	density (veh/km/ln)	travel time (seconds)
E bound	-6	8	0	**-25.0**	*-59*
W bound	142	212	62	**-51.2**	*-310*
S bound	28	0	24	**-61.9**	*-193*
N bound	0	-6	-6	**-15.6**	*-34*

Figure 4.27 Simulation results from CellNetLoad of an intersection with interchange representation and static signals compared to INTEGRATION

Even when the observed flows matched those of INTEGRATION, the density and travel time were consistently underestimated because cyclic queuing was not modelled. The average density of the westbound approach for the incident scenario was a notable exception because the queue filled both lanes, as opposed to INTEGRATION, which isolated the queue in the left lane until it spilled over and spilled back to the origin.

Explicit signal phasing was activated in CellNetLoad for the simulation results presented in Figure 4.28. The observed flows remained the same as for the static signals, except for a small change in the eastbound left-turn movement due to earlier spillback to the diverge cell. The density and travel time estimates are much improved, as shown in Figure 4.29, especially for the light flow scenario.

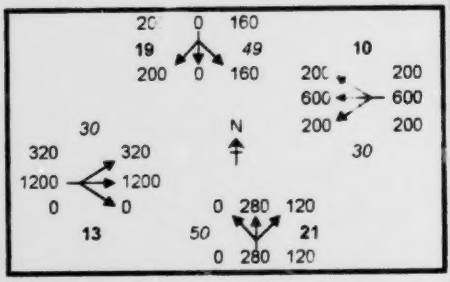
Light Flow


100 0 80
9 46
100 0 80
5
100 100
300 300
100 100
28
N
28
160 160
600 600
0 0
6
0 140 60
47 10
0 140 60

approach	demand (vph)		
	left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	160	600	0	6.3	28
W bound	100	300	100	4.9	28
S bound	80	0	100	8.7	46
N bound	0	140	60	9.7	47

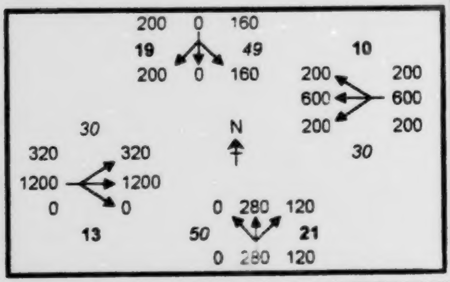
Heavy flow


200 0 160
19 49
200 0 160
10
200 200
600 600
200 200
30
N
30
320 320
1200 1200
0 0
13
0 280 120
50 21
0 280 120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	13.2	30
W bound	200	600	200	10.2	30
S bound	160	0	200	18.5	49
N bound	0	280	120	20.9	50

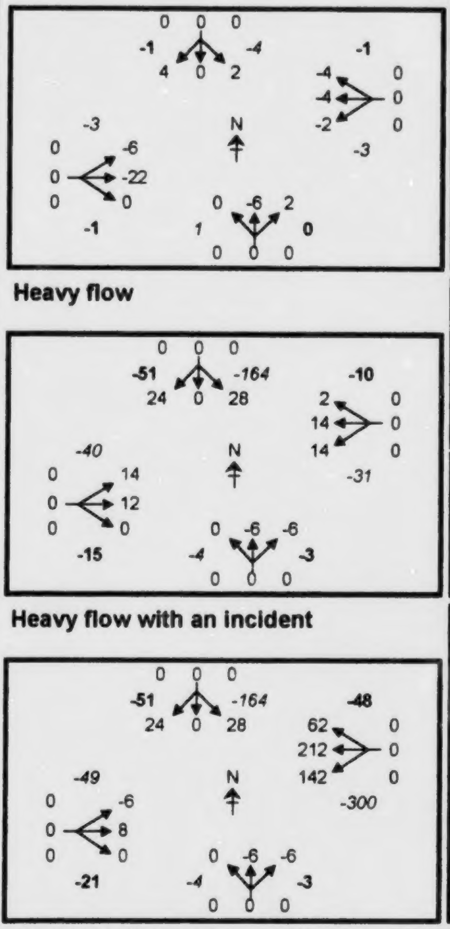
Heavy flow with an incident


200 0 160
19 49
200 0 160
10
200 200
600 600
200 200
30
N
30
320 320
1200 1200
0 0
13
0 280 120
50 21
0 280 120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	13.2	30
W bound	200	600	200	10.2	30
S bound	160	0	200	18.5	49
N bound	0	280	120	20.9	50

Figure 4.28 Simulation results from CellNetLoad of an intersection with interchange representation and signal phasing


0 0 0
-1 -4 -1
4 0 2
-4 0
-4 0
-3 N -2 0
0 -6 -3
0 -22
0 0 0 -6 2
-1 1 0
0 0 0
0 0 0
-51 -164 -10
24 0 28
2 0
14 0
-40 N 14 0
0 14 -31
0 12
0 0 0 -6 -6
-15 -4 -3
0 0 0
0 0 0
-51 -164 -48
24 0 28
62 0
212 0
-49 N 142 0
0 -6 -300
0 8
0 0 0 -6 -6
-21 -4 -3
0 0 0

Light Flow

approach	demand (vph)		
	left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	-6	-22	0	**-1.0**	*-3*
W bound	-2	-4	-4	**-0.6**	*-3*
S bound	2	0	4	**-0.7**	*-4*
N bound	0	-6	2	**-0.2**	*1*

Heavy flow

approach	demand (vph)		
	left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	14	12	0	**-14.9**	*-40*
W bound	14	14	2	**-10.5**	*-31*
S bound	28	0	24	**-50.8**	*-164*
N bound	0	-6	-6	**-3.0**	*-4*

Heavy flow with an incident

approach	demand (vph)		
	left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	-6	8	0	**-21.4**	*-49*
W bound	142	212	62	**-47.9**	*-300*
S bound	28	0	24	**-50.8**	*-164*
N bound	0	-6	-6	**-3.0**	*-4*

Figure 4.29 Simulation results from CellNetLoad of an intersection with interchange representation and signal phasing compared to INTEGRATION

4.7 Conclusions

The cell-transmission model proved to be realistic, relative to the benchmark, for networks in which the effects of lane-changing were not significant. Shockwaves were propagated correctly, as predicted by the LWR hydrodynamic model, which does not include spreading of the shockwave. The cohort FIFO algorithm respects first-in-first-out discipline among packets in the same cohort. Overtaking is permitted among packets in the same cohort only. Instabilities in the distribution of the local supply and the local demand were caused by the varying size and composition of the front cohort from which the partial demand was derived.

Several features, or lack thereof, are responsible for the error, with respect to the benchmark. Firstly, the piecewise linear equilibrium flow-density relationship prevented spreading of fans, and all uncongested flow travelled at the free speed. Shockwave speed was also slightly different from that predicted with a non-linear flow-density relationship. Secondly, the effects of lane-changing can only be included with external models, which would likely require the complexity of a microscopic simulation program or large quantities of empirical data. Therefore, asymmetric merging of lane flows due to a lane-drop and weaving upstream of diverges are not captured by the cell-transmission model. Thirdly, first-in-first-out logic is too restrictive for multi-lane diverges.

The modelling of intersections has a host of problems. The capacity reduction due to opposed movements is not modelled, and must be specified externally. Random delay, delay due to lane-changing, and delay to opposed movements is also neglected. Explicit signal phasing is necessary for the inclusion of cyclic queuing and the associated delay. Modelling spillback properly will always be a problem for the interchange-representation of intersections, especially for oversaturated movements.

Chapter 5

Extensions to Flow Propagation

5.1 Introduction

Extensions to the flow propagation logic of the traditional cell-transmission model are investigated in this chapter. These extensions are intended to improve the realism of the simulation results generated using the cell-transmission model. Extensions primarily motivated to improve the efficiency of an implementation of the cell-transmission model are beyond the scope of this study. Considerations for improvements to the intersection logic of the cell-transmission model are included in the following chapter.

The traditional cell-transmission model is capable of approximating the propagation of shockwaves as predicted by the LWR hydrodynamic model (Lighthill and Whitham, 1955; Richards, 1956). The error is associated with the discretisation of the arcs into cells, which have a length equal to the distance travelled by a vehicle during one simulation time step at the free speed. The average density of the cells defines the state of the network from which flow is predicted. As the time step is reduced, the cell length is reduced, and the average density becomes a better estimate of the true density at that location on the arc. Larger time steps reduce the execution time and memory requirements of the simulation. The upper bound of the time step is the smallest arc travel time for a vehicle travelling at the free speed for any arc in the network, since each arc must contain at least one cell. In addition to the error described above, another error is introduced by the discretisation scheme: error in representing the arc length, which is discussed in this chapter.

In the traditional cell-transmission model the flow is predicted from the density using a trapezoidal equilibrium flow-density relationship. The solution of the LWR hydrodynamic model with non-differentiable points on the flow-density relationship may contain shockwaves which violate the entropy condition, as explained in chapter 2. The impact of a non-linear equilibrium flow-density relationship on the cell-transmission model is investigated below.

The final aspect of the cell-transmission model, which will be examined, is the fluid representation of the traffic stream, which is fundamental to the LWR hydrodynamic model. The theoretical implications and some simulation results are presented for an implementation scheme which restricts density and flow to discrete quantities of vehicles.

The computational efficiency of CellNetLoad with the enhanced features is evaluated at the end of this chapter.

5.2 Non-linear equilibrium flow-density relationship

5.2.1 Implementation

The implementation of a non-linear equilibrium flow-density relationship in CellNetLoad required no modification to the program structure. The methods which return the local demand and local supply are located in the Arc object. The Cell object in question is passed as a parameter to the method, and the returned value is expressed in numbers of vehicles. A flag, which is provided as input on the command line, then stored in the Simulation class, determines if the trapezoidal relationship or the non-linear relationship is used.

The calculation of the local demand and local supply using the trapezoidal relationship is as follows.

[5.1] local demand $\tilde{Q}_D(k_i) = \tilde{q}_D(k_i)\|\mathrm{L}_a\|\frac{\Delta t}{3600}$

[5.2] local supply $\tilde{Q}_S(k_i) = \tilde{q}_S(k_i)\|\mathrm{L}_a\|\frac{\Delta t}{3600}$

[5.3] where $\tilde{q}_D(k_i) = \min(k_i v_f, q_{out,i})$,

[5.4] and $\tilde{q}_S(k_i) = \min(w(k_{\max,i} - k_i), q_{in,i})$;

the number of lanes on arc a is $\|\mathrm{L}_a\|$, the simulation time step is Δt, the arc free speed is v_f, and the arc backward wave speed w is defined in [3.1]; the cell density is $k_i = \frac{n_i}{\Delta x_i \|\mathrm{L}_a\|}$ where n_i is the cell occupancy, and Δx_i is the cell length; the maximum density, outflow capacity, and inflow capacity of cell i are given by $k_{\max,i} = k_{jam,i} f_{k,i}$, $q_{out,i} = q_{\max,i} f_{q,i} f_{s,i}$, and $q_{in,i} = q_{\max,i} f_{q,i}$, respectively; and the cell flow factor, cell signal factor and cell storage factor are $f_{q,i}$, $f_{s,i}$, and $f_{k,i}$.

[5.1]and [5.2] convert the flow values found in [5.3] and [5.4] into numbers of vehicles which are transferred during a single time step. The flow-density relationship applies to a single lane. Incidents may modify the relationship by reducing the cell flow factor $f_{q,i}$ and the cell storage factor $f_{k,i}$. A signal affects only the flow out of the cell via the cell signal factor $f_{s,i}$. The flow and signal factors may truncate the trapezoid, which does not affect the free-speed or the backward wave speed. The storage factor shifts the right side of the trapezoid without affecting the backward wave speed. However, the location of the density-at-capacity q_c is shifted, since it is defined as the midpoint of the horizontal section of the trapezoid. This, in turn, modifies the speed-at-capacity v_c.

The particular non-linear equilibrium flow-density relationship used in CellNetLoad corresponds to the non-linear speed-headway relationship used in INTEGRATION. This speed-headway relationship, given in [5.5], was developed by Van Aerde (1995).

[5.5] $$h(v) = c_1 + \frac{c_2}{v_f - v} + c_3 v$$

where constant $c_1 = c_2 c_4$,

$$\text{constant } c_2 = \frac{1}{k_{jam}(c_4 + 1/v_f)},$$

$$\text{constant } c_3 = \frac{1}{v_f}\left(-c_1 + \frac{v_c}{q_{\max}} - \frac{c_2}{v_f - v_c}\right), \text{ and}$$

$$\text{constant } c_4 = \frac{2v_c - v_f}{(v_f - v_c)^2}.$$

The corresponding flow-density relationship to [5.5] is derived by substituting $h = 1/k$ (by definition) and $v = q/k$ (according to the fundamental relationship [2.4]). The resulting equation can be manipulated algebraically into a quadratic equation, and solved for the equilibrium flow q_l as

[5.6] $$q_l(k) = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

where $a = c_3 / k$,

$$b = c_1 - 1/k - c_3 v_f \text{, and}$$

$$c = (1 - c_1 k) v_f - c_2 k \,.$$

The flow-density relationship $q_I(k)$ of INTEGRATION was applied to CellNetLoad as $\tilde{q}_I(k_i)$ with minor modifications. The density was the average cell density. The speed-flow-density parameters were set to those parameters which correspond to the particular cell, due to the presence of incidents and signals, as explained above.

The local demand and local supply for the non-linear flow-density relationship are calculated using [5.1]and [5.2], but [5.3] and [5.4] are replaced with [5.7] and [5.8].

[5.7]
$$\tilde{q}_D(k_i) = \begin{cases} q_{out} & \text{if } k_i \geq k_{c,i} \\ \tilde{q}_I(k_i) & \text{otherwise} \end{cases},$$

[5.8]
$$\tilde{q}_S(k_i) = \begin{cases} q_{in} & \text{if } k_i \leq k_{c,i} \\ \tilde{q}_I(k_i) & \text{otherwise} \end{cases}.$$

where density-at-capacity $k_{c,i} = \dfrac{1}{2}\left(\dfrac{q_{in}}{v_f} + k_{\max,i} - \dfrac{q_{in}}{w}\right)$ is derived from the definition that the density-at-capacity lies at the midpoint of the horizontal section of the trapezoidal flow-density relationship, and speed-at-capacity $v_{c,i} = q_{in} / k_{c,i}$ is used in $\tilde{q}_I(k_i)$.

5.2.2 Effect of the non-linear equilibrium flow-density relationship

The linear network with a lane drop, described in chapter 4, was used to evaluate the effect of a non-linear equilibrium flow-density relationship on the simulation results of CellNetLoad. No capacity reduction was specified for the arc upstream of the lane drop. The non-linear equilibrium flow-density relationship is shown below the density contour plot in Figure 5.1. The trapezoidal relationship is shown with a dotted line for comparison.

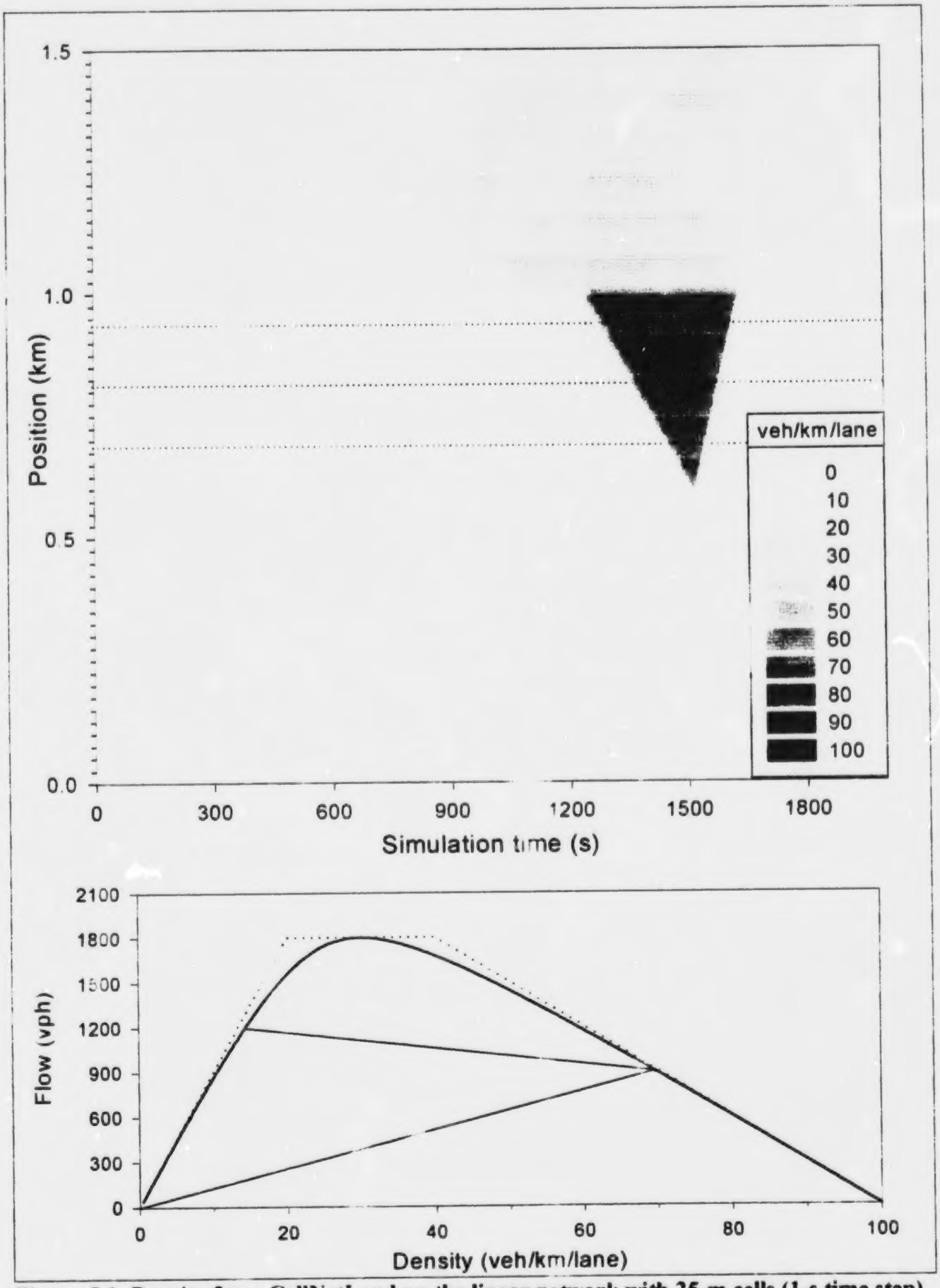


Figure 5.1 Density from CellNetLoad on the linear network with 25-m cells (1-s time step) and a non-linear equilibrium flow-density relationship

The non-linearity in the uncongested regime implies that the equilibrium speed is always smaller than the free speed. Recall that the equilibrium speed is the slope of the radius vector from the origin to the relevant point on the equilibrium flow-density relationship. Since the cell length is equal to the distance travelled at the free speed during one time step, some portion of the cell occupancy does not discharge during the time step. If it is assumed that the cell occupancy is spread evenly over the length of the cell, then the portion of the occupancy which discharges is the product of the cell density and the distance travelled at the equilibrium speed during one time step. The remaining cell occupancy causes an accumulation until the density is large enough to discharge an amount equal to the inflow.

The effect on density is opposite during congested conditions because flow is constrained by local supply instead of local demand. Speeds are still slightly smaller, but because fewer vehicles can exit the cell, fewer vehicles are permitted to enter the cell. The flow is more constrained than for a trapezoidal equilibrium flow-density relationship (henceforth, the trapezoidal case), so the density does not reach the same level.

Quantitatively, the subtle difference in density is difficult to perceive in the shading of the density in Figure 5.1. It is more evident in the lower diagram. The demand flow of 1200 vph required a density of 14.31 veh/km/lane, as opposed to the trapezoidal case which required a density of only 13.33 veh/km/lane to support a 1200-vph flow. The higher density was achieved by the occupancy accumulation explained above. The queue reached a density of 70.00 veh/km/lane in the trapezoidal case, but only 69.52 veh/km/lane in the non-linear case, due to the greater constraint on the local supply as explained above. The slower speeds increased the travel time to the intersection from 40 s to 60 s during the 1200-vph demand. So the congestion occurred slightly later.

There is also a subtle effect at the boundary between flow states. As the demand is increased during uncongested conditions, the vehicle speeds slowed. The LWR hydrodynamic model with a non-linear equilibrium flow-density relationship predicts that the edge of the upstream flow state should catch up with the rear of the downstream flow state. This smooth boundary between the flow states, called a fan, does actually appear in the non-linear case, though it was absent in the trapezoidal case.

Figure 5.2 shows the density at the three positions marked with a dotted line on the density contour plot in Figure 5.1. The slope of the step-like increases in the density represent the passing of the fan over the positions. The fan moved downstream, first crossing position 0.6875 km (dotted line), then 0.8125 km (grey line), then position 0.9375 km (black line). The fan did not spread much for the first steps because the difference in speed between the two flow states was small. However, the last step, which occurred at 1260 s, clearly shows that the fan spread wider as it travelled downstream. The slope of the step is steeper for the dotted line and is less steep for the black line.

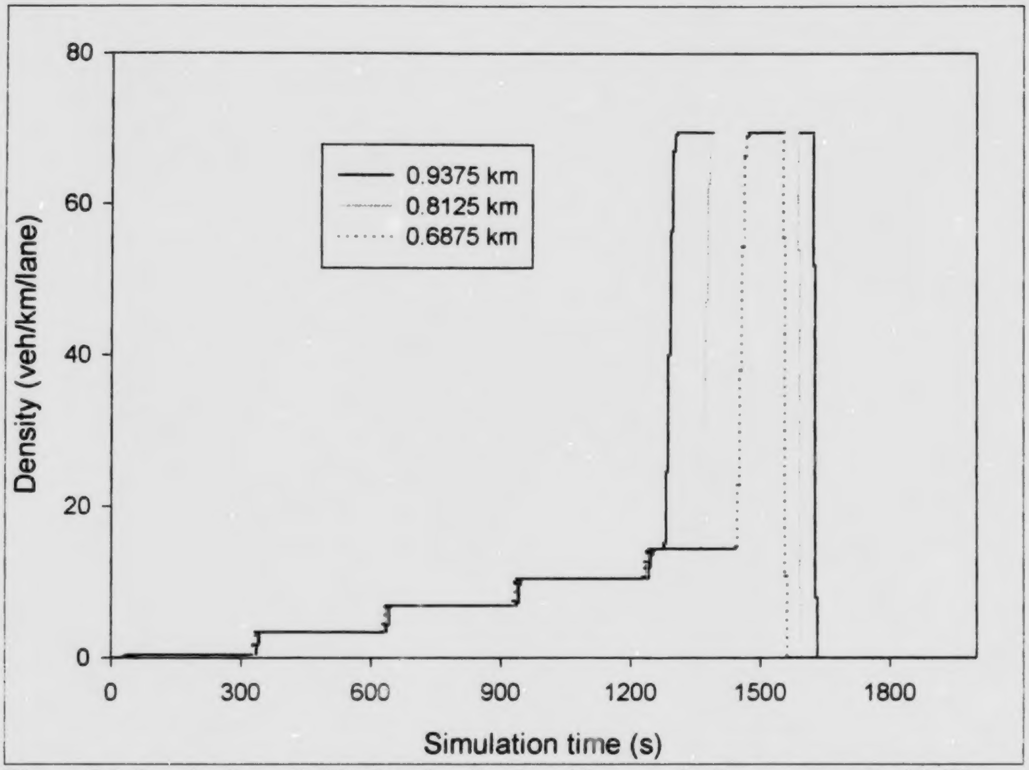


Figure 5.2 Density from CellNetLoad at fixed positions on the linear network with 25-m arcs/cells (1-s time step) and a non-linear equilibrium flow-density relationship

The shockwaves which span the congested and uncongested states are also slightly different in the non-linear case. The change in density is marginally smaller, which leads to a faster shockwave, according to the Rankine-Huginot jump condition [2.13]. This difference is obscured in Figure 5.2 by the shockwave dissipation due to numerical error. Figure 5.3 presents the density at the same positions for the network with 125-m arcs to show that the numerical error also obscures the fan. The occupancy accumulation appears as the roundness of the steps.

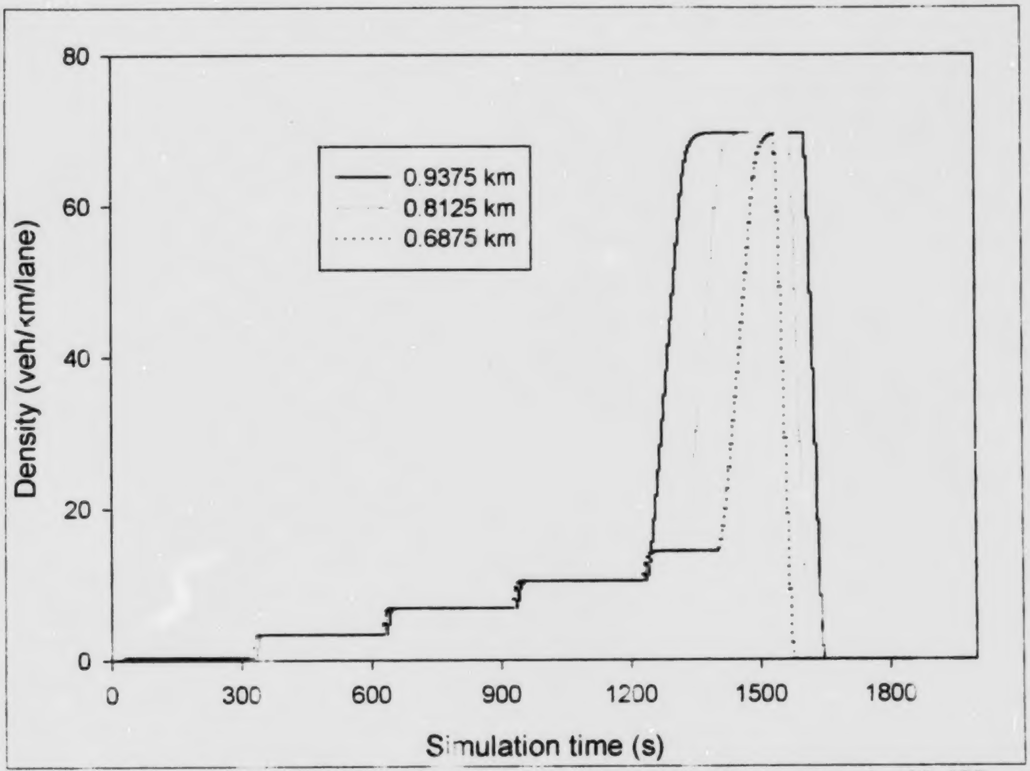


Figure 5.3 Density from CellNetLoad at fixed positions on the linear network with 125-m arcs/cells (5-s time step) and a non-linear equilibrium flow-density relationship

Thus far, the differences in the trapezoidal and non-linear cases have been negligible. However, the effect on the travel time is more significant. The estimated travel time from CellNetLoad are shown in Figure 5.4. The steps in Figure 5.4 show the increased travel time in the uncongested state, which is consistent with the estimates from INTEGRATION. Recall from chapter 4, that the trapezoidal case predicted the same travel time for all flows in the uncongested regime. (The unexpected increase in travel time for larger time steps during the first 300 s is a result of truncation error due to the implementation of the fluid representation of vehicles in CellNetLoad.) The roundness of the steps is evidence of the occupancy accumulation, which is less pronounced for larger time steps. The path travel time peaked at 170 s, which is 10 s higher than the trapezoidal case. The hump in the path travel time at the end of the congestion is due to a delayed clearance. The occupancy is redistributed over the length of the entire cell at each time step, which prevents some portion of the occupancy from clearing during the next time step. This delayed clearance actually counteracted the underestimation of path travel time at the end of the congestion by providing a few extra vehicle probes beyond the end of demand at 1500 s.

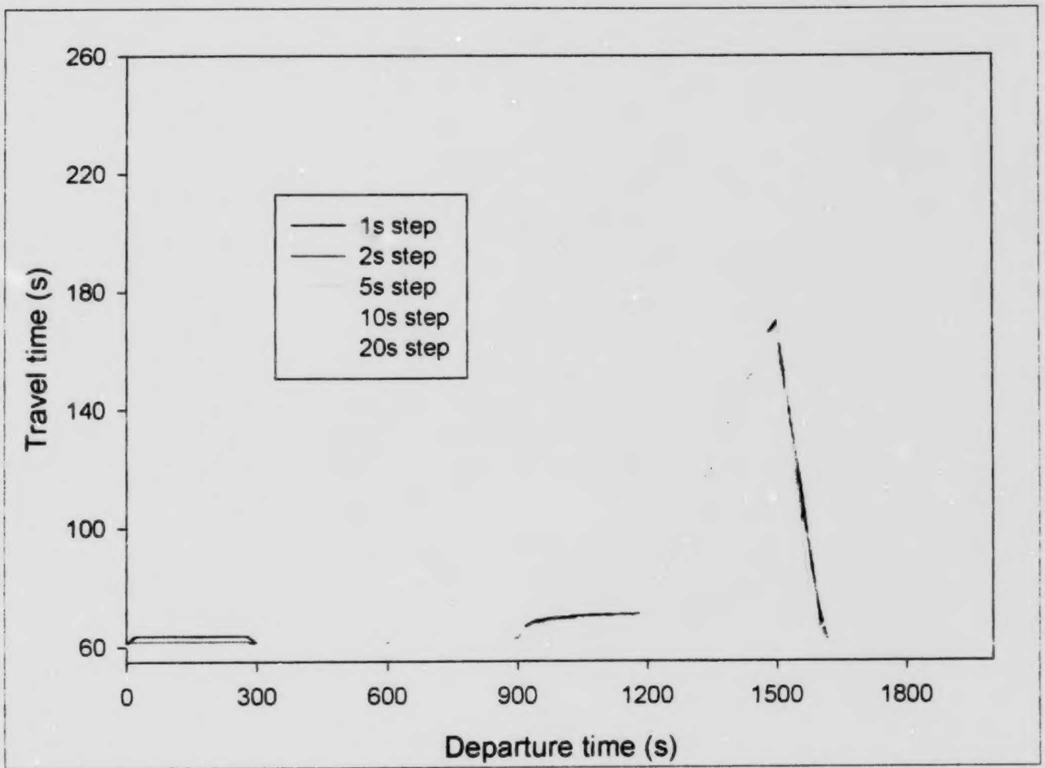


Figure 5.4 Path travel times from CellNetLoad on the linear network with a non-linear equilibrium flow-density relationship

In summary, the effect of the non-linear equilibrium flow-density relationship on the observed density and shockwave speeds is small. Although the non-linear equilibrium flow-density relationship increased the numerical error in propagating shockwaves and fans, at least it permitted the modelling of fans. Larger fans appear during recovery from congestion. Although shockwaves may appear to be fans during the trapezoidal case due to numerical error, this error shrinks with smaller time steps, so the trapezoidal case does not properly model fans. The non-linear case also models the expected increase in travel time with increasing demand during uncongested conditions. The clearance delay does not seem to be a drawback since it reduces the underestimation of travel times caused by a lack of probes.

5.3 Arc discretisation

The cell-transmission model represents the arc length using an integer number of cells, each with a length determined by the distance travelled during one time step at the free speed of that arc. Because each arc must contain at least one cell, the time step is bounded above by the arc with the shortest travel time at the free speed.

5.3.1 Error in arc length

Since all arcs in the network are discretised according to the same time step, the arc length may be modelled with an error that is smaller than one cell length. This error in length affects the travel time for the arc and the storage capacity. For large arcs, the relative error is small and acceptable. Even for short arcs, where the relative error of the arc travel time is large, the relative error in *path* travel time for a path which contains that arc is small and excusable. The underestimated travel time may have a more severe effect in networks with co-ordinated signal timings, but the platoons tend to arrive early, so the green phase is not wasted.

A more serious concern for modelling insufficient arc length is the underestimated storage capacity. On short arcs this inadequate storage may lead to premature spillback, which can have drastic effects on the upstream flows. CellNetLoad compensates for this fault by allowing for the last cell of all arcs to have an increased jam density without changing the cell length, free speed, or backward wave speed. This feature is activated using a flag on the command line. The flag alerts the simulation to initialise the storage factor $f_{k,i}$ for the last cell i to

$$f_{k,i} = \frac{\Delta x_i + \bar{x} \bmod \Delta x_i}{\Delta x_i}$$

where the cell length $\Delta x_i = v_f \Delta t$, the free speed is v_f, the time step is Δt, and the arc length is $\bar{x}$.

The effect of the storage factor is the same as an increase to the jam density, which stretches the middle section of the trapezoidal flow-density relationship, and stretches the entire non-linear curve. This stretching causes shockwaves to be propagated more slowly, which is actually more realistic since the cell is shorter. In the trapezoidal case only shockwaves which span the uncongested and congested regimes will be propagated at slower speeds.

The effect of more realistic spillback would be evident on a network where an off-ramp is signalised. An error in the length of the off-ramp would cause premature spillback on the freeway. No test networks were developed to investigate the quantitative impact of this premature spillback, but a more complex network is presented in chapter 7 to demonstrate this effect and the error in arc travel times.

An alternative solution to the inaccurate arc length is to increase the length of the last cell. However, a cell that is longer than the maximum distance travelled during one time step would make clearance of a cell impossible, because the remaining occupancy would be spread evenly over the length of the cell after each time step. This phenomenon also occurs on cells of the regular length in the non-linear case, as explained in the previous section. The clearance of cells of regular length is not a significant concern in the non-linear case because the equilibrium speed approaches the free-speed for smaller densities. This difficulty in cell clearance would be more severe for longer cells.

5.3.2 Reduction in time step due to intersection model

The modelling of intersections further constrains the time step. If an arc happens to have an upstream node with multiple arcs in the backward star, then the first (and only) cell must be a merge cell. If the downstream node has multiple arcs in the forward star, then the last (and only) cell must be a diverge cell. Since a cell cannot be a merge and a diverge cell simultaneously in the traditional cell-transmission model, then at least two cells are required to represent this arc. This effectively halves the upper bound for the time step, and may quadruple the simulation execution time.

If the downstream node has multiple arcs in the forward star and the backward star, then three cells are required for this arc: a merge cell, a diverge cell, and an ordinary cell to carry the traffic stream to the next arc in the interchange representation of the intersection. The minimum of three cells will reduce the upper bound for the time step by two thirds, and may increase the simulation execution time by a factor of nine.

This phenomenon is not present in CellNetLoad because of the implementation of a generalised intersection logic which permits one cell to act as a merge cell and a diverge cell. Also, no ordinary cell is required as a buffer between a diverge cell and a merge cell. So every arc can be modelled using a single cell.

The generalised intersection model is iterative, hence computationally less efficient than the traditional merge and diverge logic. The savings in execution time, due to the permission of larger time steps, may result in a net savings in computational requirements. This topic is reserved for future research.

The generalised intersection model limits the level of potential parallelism in the implementation, since the exit flow from all cells of the arcs in the backward star of a node are interdependent.

More details of the intersection model are given in chapter 6. It is hoped that the improvements in realism of the simulation results offset these drawbacks.

5.3.3 Variable time step

Allowing a different time step for each arc has many advantages, and some major complications. Although no such feature was implemented or tested in CellNetLoad, a brief discussion of this feature is worthwhile.

The advantages primarily involve some savings in simulation execution speeds, though a secondary consideration is that these savings may not come solely at the expense of accuracy, as in the traditional cell-transmission model. The upper bound for the time step could be arc-specific. For example, long freeway arcs could be modelled with long cells even if the off-ramp is very short. In fact, the cell length could be customised for each arc, such that the arc length is modelled exactly. This also permits different facilities to be modelled with different accuracy, though which facilities would benefit most from short cells is unclear. Small time steps for freeways could reduce numerical error in shockwaves and fans, whereas small time steps for surface streets could allow improvements in intersection modelling. This feature also permits more flexibility in customising the execution speed, which is important for real-time applications.

An implementation of the cell-transmission model with variable arc discretisation would be consistent with the theory for the propagation of flow on each arc. The complication is the transfer of flow from one arc to the next. One possibility for such an implementation involves the use of buffers at each end of each arc. These buffers would interpolate the local demand or local supply when queried between time steps by adjacent arcs. The impact of these buffers on shockwaves and fans is unknown. Another complication is the collection of output statistics since arc outflow may not coincide with arc inflow across nodes. Other complications surely exist which cannot be predicted before trying to implement this feature.

Ziliaskopoulos and Lee (1997) implemented a cell-transmission model which represented arterials with short cells and freeways with long cells. The time step on arterials was 2 seconds, which permitted some rudimentary gap acceptance at intersections. The time step on freeway arcs was a power of 2. Little detail on the rest of the implementation is given.

It was decided not to investigate this feature here because it would greatly complicate the data structures and intersection model of CellNetLoad, perhaps so much that it would be difficult to test the other enhancements without interference from the buffers. In addition, this feature is primarily motivated by a savings in execution time, which is not the focus of this research.

5.4 Discretisation of flow

The cell-transmission model borrows the fluid representation of traffic flow from the LWR hydrodynamic model. The fluid is represented using real numbers in analytical calculations. An interesting challenge of implementing the cell-transmission model in a computer simulation is the representation of traffic as a fluid. Computers represent real numbers with limited precision, so round-off error may accumulate throughout the simulation, making fractions of vehicles seem to appear and disappear.

In reality, vehicle quantities are integers. This section explores the quantitative impact on the simulation results of the integration of discrete vehicles with the cell-transmission model. An additional potential benefit is the improvement of simulation execution times, since the time spent moving very small amounts of flow would be saved.

5.4.1 Implementation

Care was taken during the implementation of CellNetLoad to respect the flow conservation law strictly. This was accomplished by truncating flows, using bucket rounding to split the packets of cohorts, and correcting round-off errors. The float data type, which was used to store all non-integer numbers, had reliable precision in the sixth decimal place. The correction of round-off errors was made in the fifth decimal place as a precaution against their accumulation. The double data type was available if more precision was desired, but it seemed to be an unnecessary consumption of memory.

The size of the vehicle unit is not fixed. A vehicle unit larger than 10^{-5} must be specified on the command line. Each packet in the simulation is an integer multiple of the vehicle unit. A vehicle unit equal to 0.001 was chosen as the default for all testing. The demand cohort at an origin may contain packets with fractional vehicle units. These fractions accumulate in the latent demand cohort each time step until they are sufficient to load a whole vehicle unit into the last cell of the origin. Similarly, if the calculation of the flow finds that a fractional number of vehicle units should leave a cell, then the decimal portion of that fractional flow is not used. This truncation error is acceptable for large flows and small vehicle units, but if a larger vehicle unit is used, then the truncation error may become large.

An enhanced feature of CellNetLoad is the probabilistic rounding of flow to whole vehicle units. This feature is activated using a flag on the command line. The implementation of this feature required modifications to the program logic which prepares vehicles to leave the cell, and which loads the demand at the origin.

The modification concerning the preparation of vehicles to leave a cell involved altering the immediate truncation of the calculated flow. Once the flow calculation has been performed, the value is rounded up to the nearest vehicle unit with probability equal to the decimal quantity of a vehicle unit. This is accomplished by comparing the decimal value to a randomly generated value between 0 and 1. If the decimal exceeds the random number, then the flow is rounded up. Otherwise, the flow is rounded down.

For example, based on the local demand and the local supply, it was found that 4.36 vehicle units should exit the cell during the upcoming time step. Then there is a 36 percent chance that 5 vehicle units will be moved from the cohort queue to the cohort list of vehicles ready to depart the cell, and there is a 64% chance that only 4 vehicles will be moved.

The situation in which more flow is to be moved than is present in the cell could never occur. Each packet is an integer multiple of the vehicle unit, so the occupancy of the cell must also be an integer multiple of the vehicle unit. The local demand is bounded from above by the occupancy of the cell, and the calculated flow is constrained by the local demand. Since the calculated flow is bounded from above by the occupancy of the cell, and the occupancy is an integer multiple of the vehicle unit, there is no chance that the calculated flow could be rounded up so that it exceeds the cell occupancy.

The modification concerning the loading of demand at the origin also involved altering the truncation. Packets with a fractional number of vehicle units are cloned from the demand cohort and placed in the latent demand cohort at the beginning of each clock tick event. Then an integer number of vehicle units is moved from each packet in the latent demand cohort to the cohort queue of the last cell of the origin. The number of vehicle units moved from each packet is the fractional size of the packet rounded up with probability equal to the decimal quantity of a vehicle unit in the packet, and it is rounded down otherwise. It is possible that the number of vehicle units to be removed from a packet in the latent demand cohort exceeds the size of the packet. In such a situation, a packet with a negative size remains in the latent demand cohort. During the next clock tick event, this packet is added to the new cloned packet from the demand cohort.

5.4.2 Effect of the discretisation of flow

This feature was tested using the same linear network with a lane drop and a non-linear flow-density relationship. The vehicle unit was set to 1.0, and the probabilistic rounding feature was activated. A variety of time steps were tested. The density contour plot for a 5-s time step is shown in Figure 5.5.

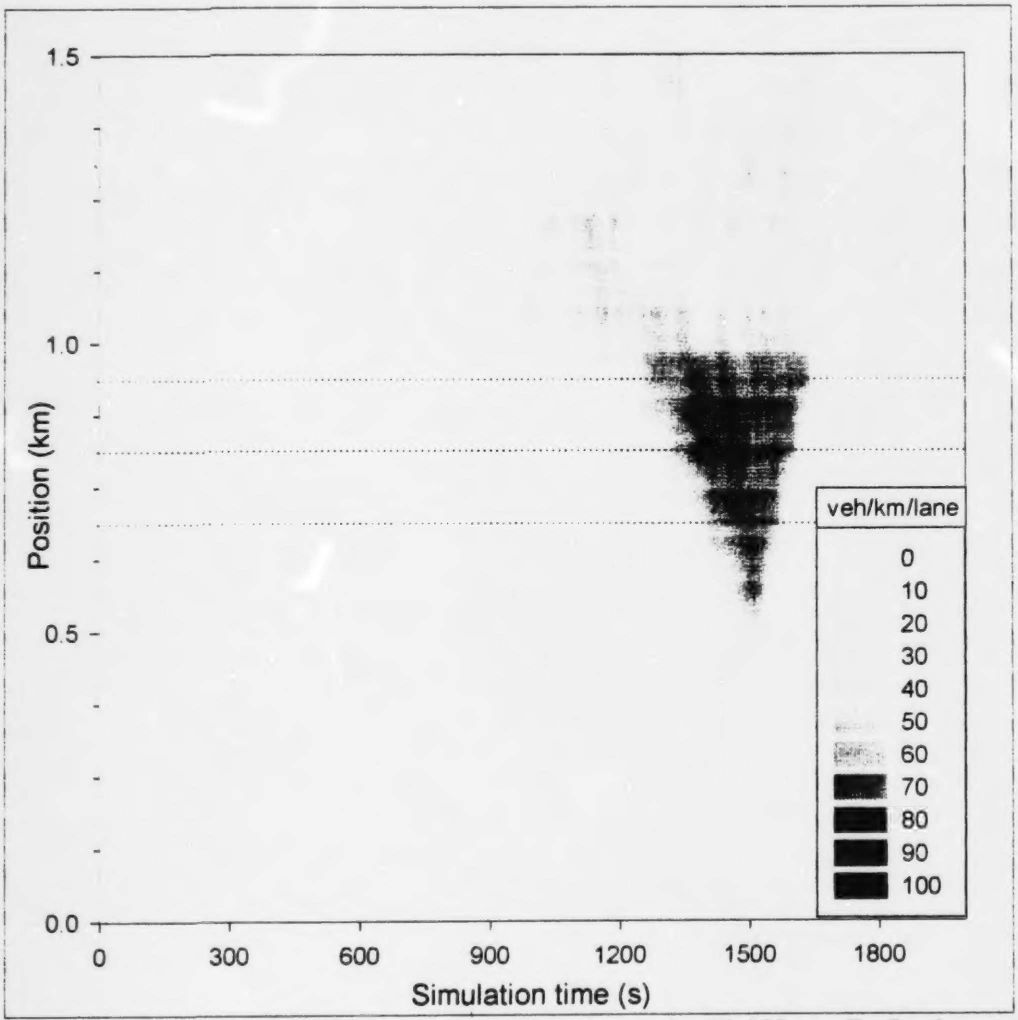


Figure 5.5 Density from CellNetLoad on the linear network with 125-m cells (5-s time step) and probabilistic rounding to integer vehicle units

The most striking effect of the probabilistic rounding is the presence of darker streaks of higher density during levels of demand that should dictate uncongested conditions. This results from a bias in the rounding logic which produces a net reduction in flow. The origin demand loading logic is unbiased because packets with negative size may occur. However, as explained above, the rounding logic will never send more vehicles than occupy a cell, but sometimes a portion of the calculated flow may not be used. This bias is even more pronounced for the truncation logic used as the basic logic of CellNetLoad. The bias becomes more prominent with larger vehicle units and smaller time steps. During low flow, the bias creates temporary increases in the cell density, which is likely discharged in the next few time steps. As the flow approaches the capacity, the recovery from these temporary increases in cell density is less likely.

In the example presented in Figure 5.5, the product of the time step of 5 s and the capacity flow of 1800 vphpl yields a maximum discharge of 5 vehicles (or vehicle units) from a cell on a 2-lane arc, and 2.5 vehicles from a cell on a 1-lane arc. For each demand of 60, 600, 1200, and 1800 vph the expected flows are 0.083, 0.833, 1.667, and 2.5 vehicles per time step, which correspond to one vehicle unit discharged each 12.0, 1.2, 0.6, and 0.4 time steps. The streaks during low flow are the result of occasional single-vehicle discharges which propagate downstream at their expected speeds. The streaks during high flow are the result of the occasional discharge of a flow that exceeds capacity, but is then partially held back in the next cell. This penalisation of high flows tends to behave like a capacity reduction, which is seen in Figure 5.5 when the 1800-vph demand from 900 to 1200 s reaches the single-lane arc. Unfortunately, this capacity reduction tends to occur on the single-lane arc, and not upstream of the lane drop as is the case in INTEGRATION due to lane changing.

The queue growth is stochastic, so the deterministic, step-like increases in the queue shown in Figure 4.9 do not appear in Figure 5.5. The queue grows over 100 m longer than for the CellNetLoad simulation with truncation after 0.001 vehicles (Figure 5.1) due to the capacity reduction.

Another perspective is offered by the plots of the density at fixed positions shown in Figure 5.6. The positions are marked with dotted lines on the density contour plot in Figure 5.5. The equivalent plot for the CellNetLoad simulation with truncation after 0.001 vehicles is displayed in Figure 5.3.

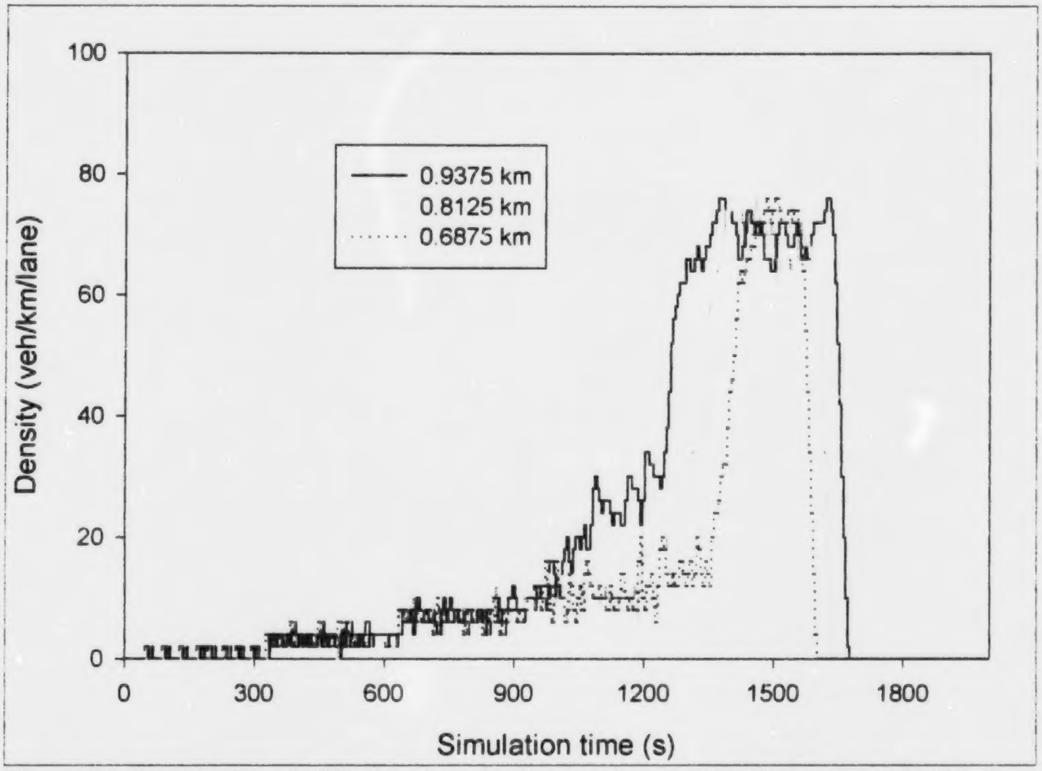


Figure 5.6 Density from CellNetLoad on the linear network with 125-m arcs/cells (5-s time step) and probabilistic rounding to integer vehicle units

The trajectories of individual vehicles during the 60-vph demand from 0 to 300 s is reminiscent of the INTEGRATION plots in Figure 4.5, though the departure headways are not constant in CellNetLoad. The fact that the departure headways are not constant creates jumps in density from 300 to 600 s because one cell may contain 0, 1, or 2 vehicles. Note that the spikes of high density from 600 to 900 s always follow troughs of low density. This is evidence of the recovery process after a flow restriction. The black, solid line during the 1800-vph demand from 900 to 1200 s shows that some queuing occurred upstream of the lane drop due to the capacity reduction. Though the queue recovered twice, it eventually grew to occupy about half of the cell. The oversaturation demand from 1200 to 1500 s caused a queue to grow past all three positions earlier, and not deterministically as it did in Figure 5.3. The clearance of the queue occurred at the same speed, and with the same spreading of the shockwave as in Figure 5.3, but it was later due to the longer queue.

The discretisation of flow would seem to be a promising mechanism for modelling the capacity reduction found in INTEGRATION within the constructs of the cell-transmission model. This would be a false conclusion. The mechanism is entirely different, therefore it seems to be appropriate only under particular circumstances. Figure 5.7 is included to show how poorly this mechanism can model the capacity reduction. The small time step forces the bias in the logic to greatly overestimate the capacity reduction. The queue grows as the 1200-vph demand reaches the lane drop.

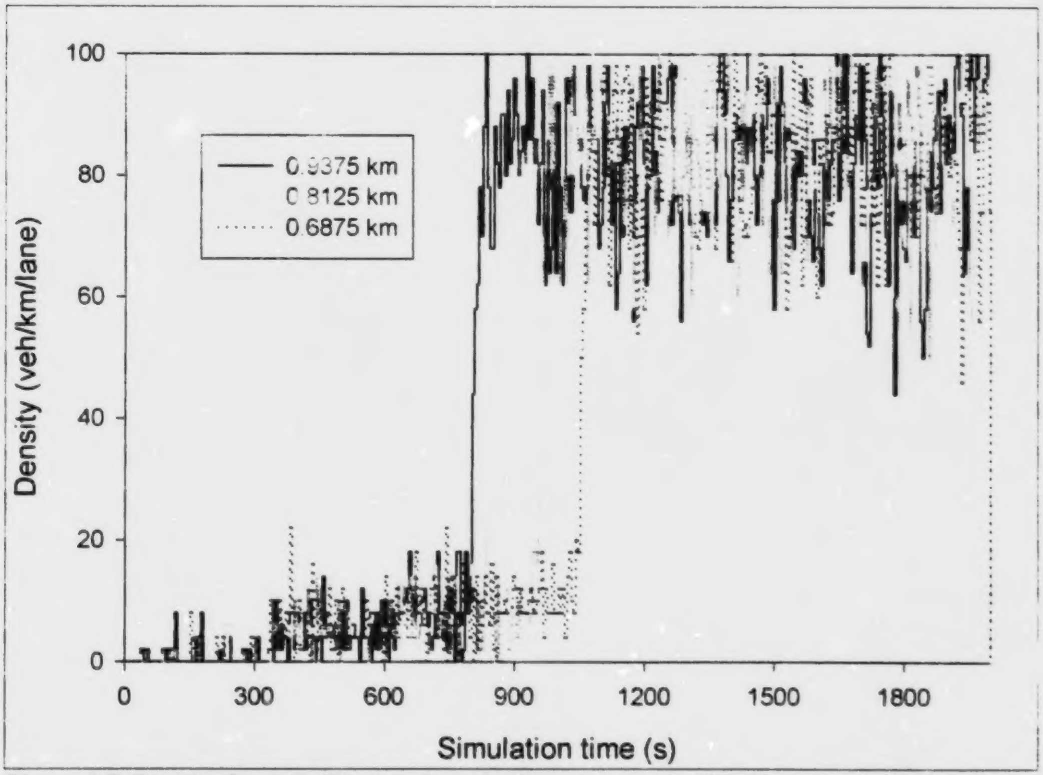


Figure 5.7 Density from CellNetLoad on the linear network with 25-m arcs/cells (1-s time step) and probabilistic rounding to integer vehicle units

The path travel times are presented for all of the tested time steps in Figure 5.8. For uncongested flow and large time steps the only difference from Figure 5.4 is the stochasticity in the flow, which also appears in INTEGRATION in Figure 4.7. The flow restraint mimics the capacity reduction found in the INTEGRATION simulation, causing increases to the path travel time for all time steps. (There was no externally applied capacity reduction specified in the CellNetLoad simulation.) Coincidentally, the path travel times for the 10-s, 5-s and 2-s time steps, which have cell lengths of 250 m, 125 m, and 50 m, seem to match the path travel times found in INTEGRATION for the simulations with 125-m, 50-m, and 25-m arcs, respectively. The cell lengths in CellNetLoad do not match the arc lengths in INTEGRATION. The difference in the mechanisms and their resulting impacts is obvious when examining the CellNetLoad results for the 1-s time step. The capacity reduction is far larger than predicted by INTEGRATION.

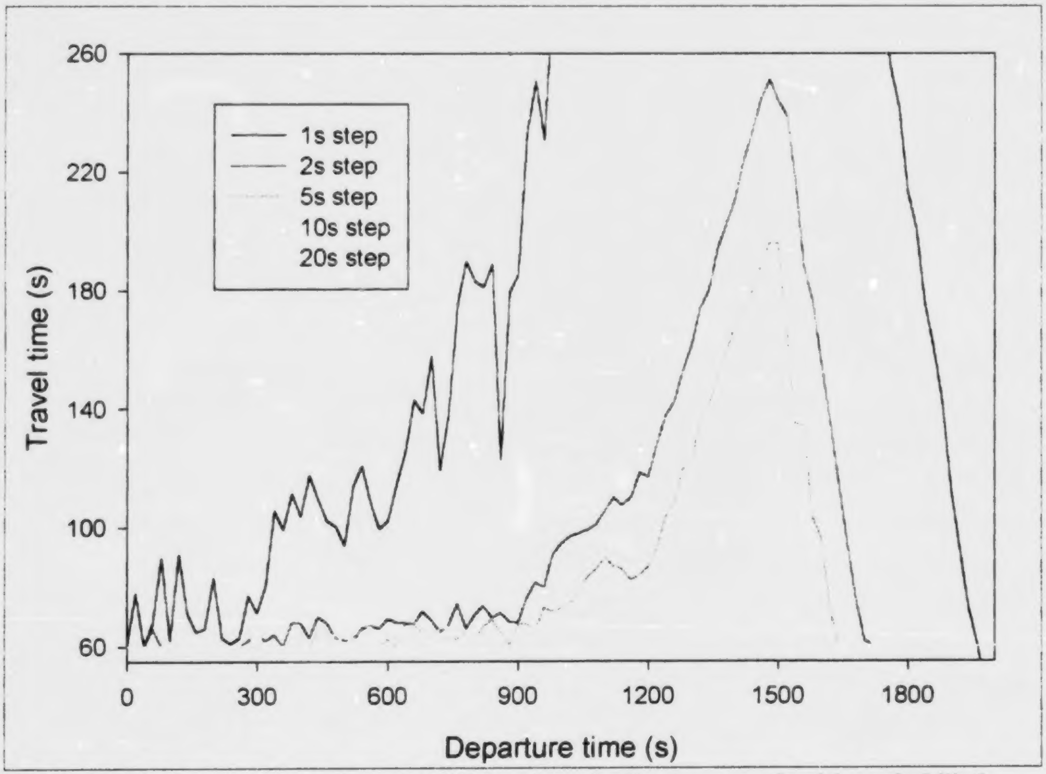


Figure 5.8 Path travel times from CellNetLoad on the linear network with probabilistic rounding to integer vehicle units

Another important drawback of the discretisation of flow is that the model will no longer converge to the solution of the LWR hydrodynamic model. So determining an appropriate time step becomes a difficult problem. Nevertheless, for large time steps, this feature promises to reduce simulation times without distorting the simulation results.

5.3 Computational efficiency

If the solution to the dynamic network loading problem is used in a real-time environment for route guidance, then the execution time of the cell-transmission model may be even more important that the accuracy of the travel time estimates. The quantitative evaluation of the computational efficiency of CellNetLoad is not the primary objective of this research. Nevertheless, it may be interesting to determine the relative cost, in terms of execution time, of the proposed enhancements to the cell-transmission model.

The program execution time was broken down by initialisation, input-processing, simulation, and post-processing of the output. The simulation time is representative of the performance of the cell-transmission model, whereas the initialisation, input-processing, and post-processing are specific to the input-output structure of the implementation. CellNetLoad reports the execution times for each of the above components of the program execution, as well as the cumulative execution time of each type of event. Unfortunately, the execution times were determined using the clock time, because the process time is not available in the chosen development environment. The operating system introduced inconsistencies in the evaluation of the execution times, which would have been avoided if the process time was available.

5.3.1 Analysis of complexity

The linear network was used to determine the coefficients for the simulation execution time. The demand was set to 300 vph throughout the simulation in order to avoid complications introduced by congestion, and to reduce the number of demand events. The same simulation was performed ten times for each time step. Five time steps were used: 20, 10, 5, 2, and 1 s. For each simulation run, the execution time of the simulation was recorded, along with the cumulative execution time of the clock tick events and the output statistics events. Three regression analyses were performed to determine the coefficients of the simulation execution time.

The first regression was used to determine the execution time of the clock tick events. The clock tick event is performed once per time step. So, the number of clock tick events executed is equal to $\bar{t}/\Delta t$, where $\bar{t}$ is the duration of the simulation and Δt is the time step. The duration of the simulation was specified as 2000 s.

During each time step, the flow is calculated and moved from each cell. Since there are no intersections—an intersection is a node with multiple arcs in the backward star and/or the forward star—and no congestion in the linear network, the time spent at each cell should be equal. The total number of cells is equal to $\sum_{\mathrm{a}\in\mathrm{A}} \left\lfloor \frac{\bar{x}_{\mathrm{a}} \quad v_{f,\mathrm{a}}}{\Delta t} \right\rfloor + \|\mathrm{O}\|$, where A and O are the sets of arcs and origins, and $\bar{x}_{\mathrm{a}}$ and $v_{f,\mathrm{a}}$ are the length and free speed of arc a. The arc lengths were modified for each time step so that each arc contained exactly one cell, which makes the $\lfloor \; \rfloor$ operator unnecessary.

The regression equation for the cumulative execution time of the clock tick events has the form

$$c_1 + c_2\left(\bar{t}/\Delta t\right) + c_3\left(1 + \sum_{\mathrm{a}\in\mathrm{A}} \frac{\bar{x}_a \quad v_{f,\mathrm{a}}}{\Delta t}\right)\left(\bar{t}/\Delta t\right),$$

which can be rearranged to the quadratic form

$$c_1 + (c_2 + c_3)(\bar{t}/\Delta t) + \frac{c_3}{\bar{t}}\left(\sum_{a\in A} \bar{x}_a / v_{f,a}\right)(\bar{t}/\Delta t)^2 .$$

Constraints were specified for the regression $c_1, c_2, c_3 \geq 0$. The regression found that the coefficients had values of 0.2320 s, 0 s/step, and 5.803×10^{-5} s/cell with an r^2 of 0.9984. The curve is shown in Figure 5.9 with error bars two standard deviations long.

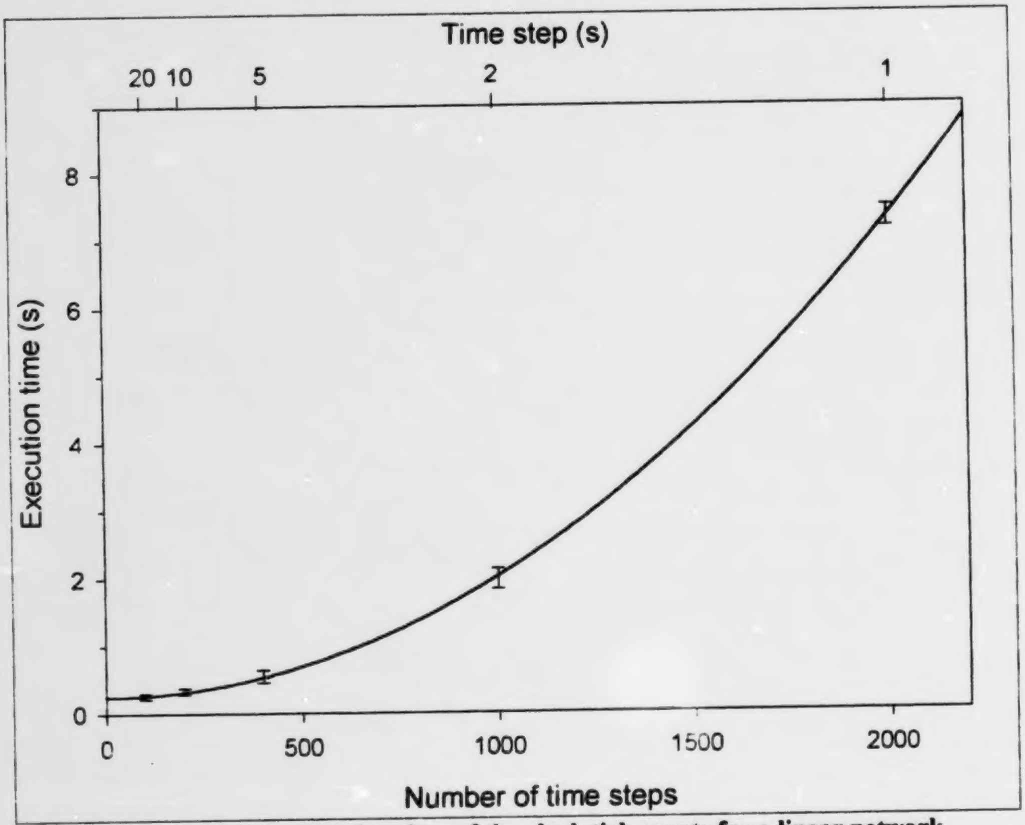


Figure 5.9 Cumulative execution time of the clock tick events for a linear network

The second regression was used to determine the execution time of the output statistics events. The output statistics events is performed once per output interval, which was made equal to the time step for these simulations, and again at the end of the simulation. During each output statistics event, the statistics are printed for each arc and each origin. So, the regression equation has the form

$$c_1 + c_2\left(1 + \bar{t}/\Delta t\right) + c_3\left(1 + \sum_{a\in A} \frac{\bar{x}_a \quad v_{f.a}}{\Delta t}\right)\left(1 + \bar{t}/\Delta t\right),$$

which can be rearranged to the quadratic form

$$c_1 + c_2 + c_3 + \left(c_2 + \frac{c_3}{\bar{t}}\left(1 + \sum_{a\in A} \bar{x}_a / v_{f.a}\right)\right)\left(\bar{t}/\Delta t\right) + \frac{c_3}{\bar{t}}\left(\sum_{a\in A} \bar{x}_a / v_{f.a}\right)\left(\bar{t}/\Delta t\right)^2.$$

Constraints were specified for the regression $c_1, c_2, c_3 \geq 0$. The regression found that the coefficients had values of 0 s, 0 s/interval, and 2.579×10^{-4} s/arc (or origin) with an r^2 of 0.9992. The curve is shown in Figure 5.10 with error bars two standard deviations long.

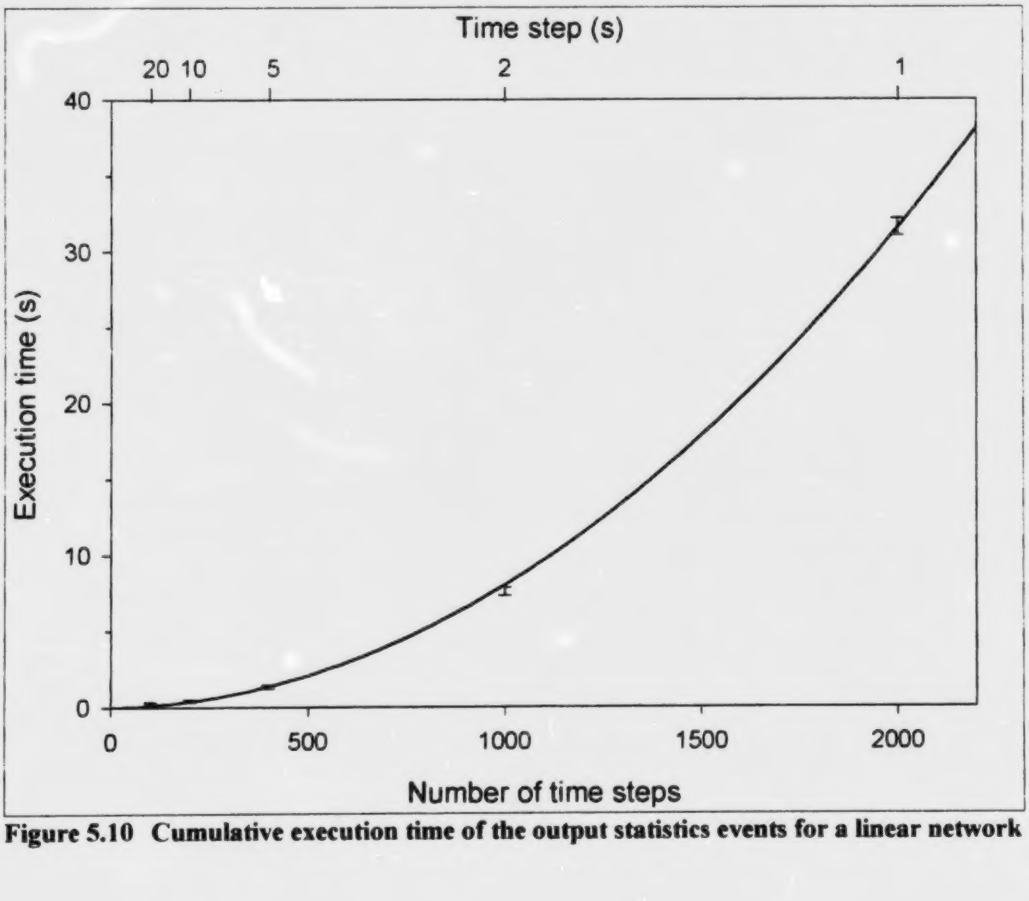


Figure 5.10 Cumulative execution time of the output statistics events for a linear network

The third regression was used to determine the remainder of the simulation execution time, which consisted of the event manager, and the other types of events. This simulation had one demand event, one route event, and one simulation terminate event. The same form of the regression equation was used

$$c_1 + c_2(\bar{t}/\Delta t) + c_3(\bar{t}/\Delta t)^2 .$$

The coefficients were found to be 0.079, 7.55×10^{-5}, and 2.61×10^{-8} with an r^2 of 0.549. The curve is shown in Figure 5.11 with error bars two standard deviations long. The small r^2 is a consequence of the high degree of variability in the data.

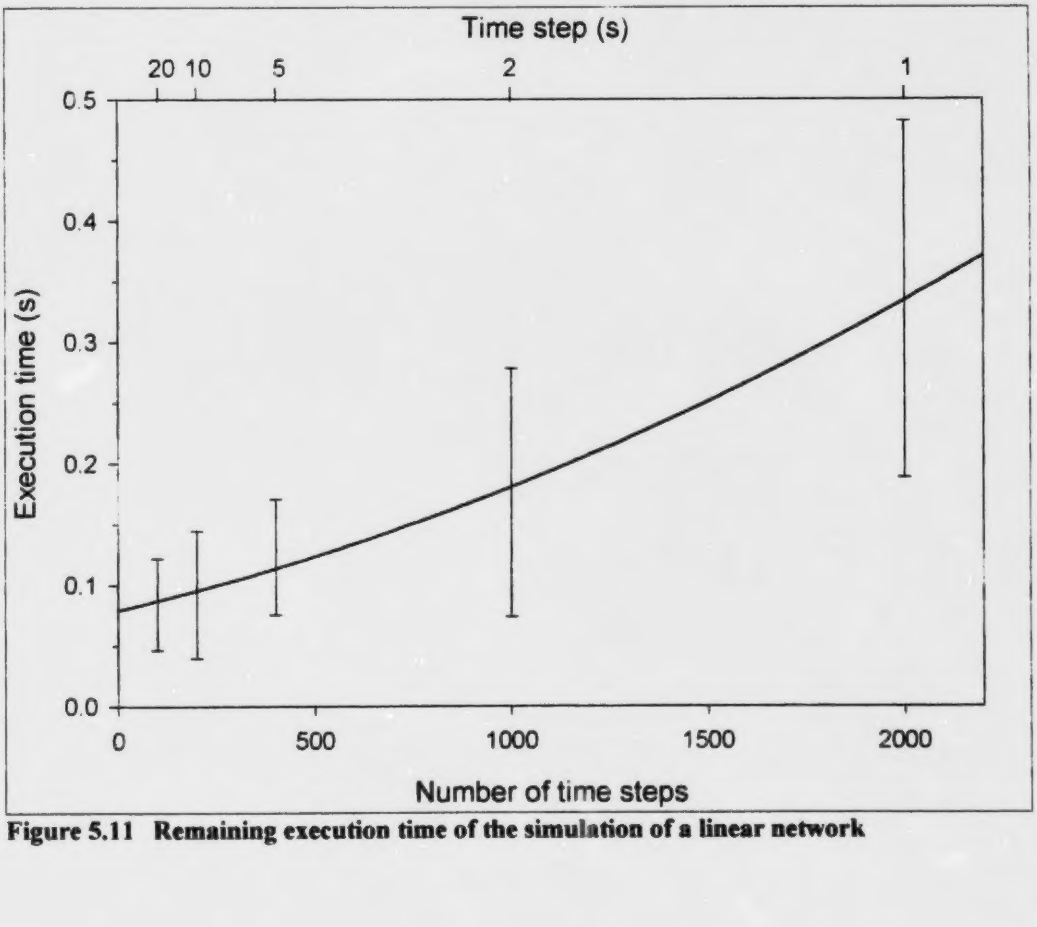


Figure 5.11 Remaining execution time of the simulation of a linear network

5.3.2 Comparison of execution times for active enhancements in CellNetLoad

The regression coefficients were used to develop a prediction of the simulation execution time for the linear network used in testing of CellNetLoad. The output interval was equal to the time step, except for the 1- and 2-s time steps which used an output interval of 5 s. The predicted simulation execution time is shown with a solid line on Figure 5.12 and Figure 5.13

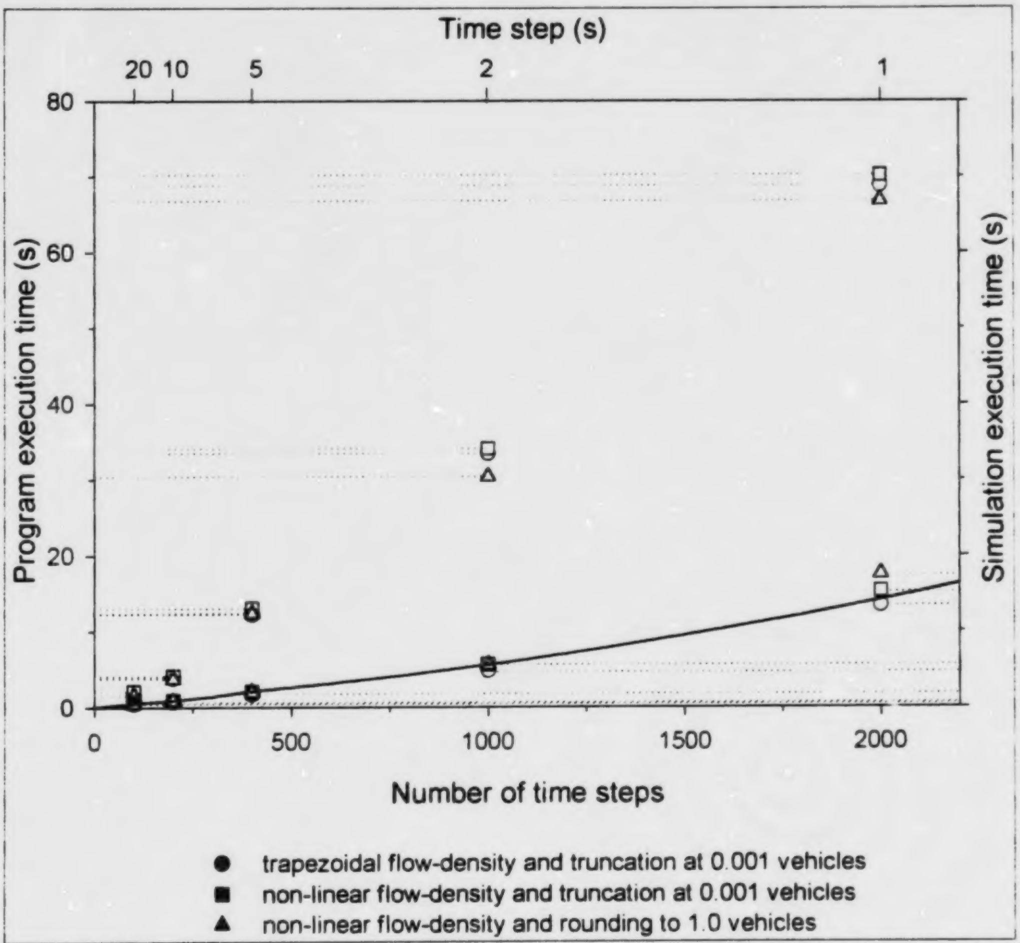


Figure 5.12 Execution time and simulation time of CellNetLoad on the linear network

The quantitative comparison of the execution times of CellNetLoad is shown in Figure 5.12. Each symbol represents one run of CellNetLoad. The simulation component of the execution time is shown with unfilled symbols. The whole execution time is shown with symbols filled in grey. The vertical axes are displayed with linear-scaling. The execution time and the simulation time are of $O\left((\bar{t}/\Delta t)^2\right)$.

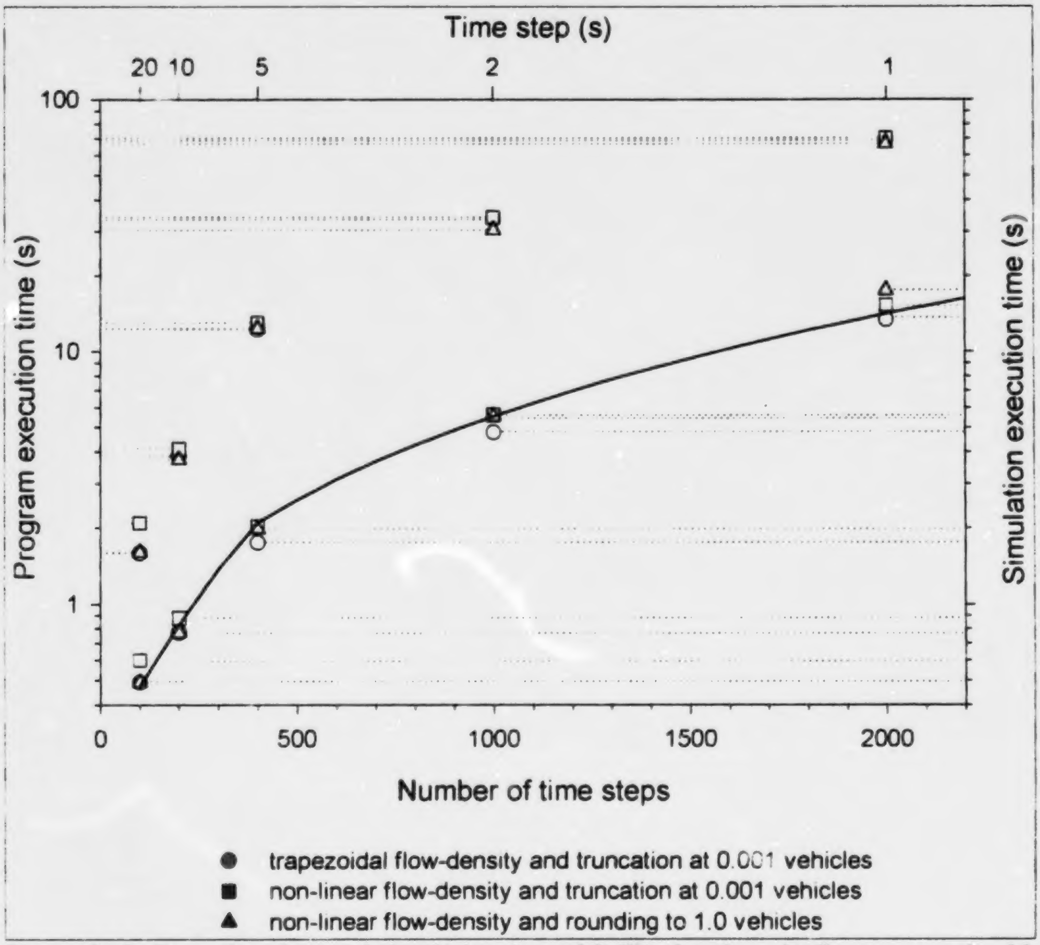


Figure 5.13 Execution time and simulation time of CellNetLoad on the linear network with log-scaling

The same data is plotted in Figure 5.13 with log-scaling of the vertical axes to better compare the performance of the basic and enhanced simulations The non-linear equilibrium flow-density relationship consistently increases the simulation time compared to the trapezoidal case. For large time steps the probabilistic rounding to the nearest integer number of vehicles tends to reduce execution time, due to the savings in moving small packets. For small time steps the excessive capacity reduction caused by the bias of the rounding tends to increase the execution time.

5.4 Conclusions

Equilibrium flow-density relationships are derived from observations of traffic that is rarely in an equilibrium state. Any flow-density relationship is considered to be approximate. The choice of the functional form, which is then fit to the observed data using regression, is rather arbitrary, provided that there is an increasing flow as the uncongested density increases, and there is a decreasing flow as the congested density increases.

The use of a non-linear equilibrium flow-density relationship with the cell-transmission model poses no significant problems. Compared to a trapezoidal functional form, the non-linear case has several advantages. It permits the modelling of fans, and it predicts an increased travel time with increasing uncongested flow. The disadvantage is that shockwaves dividing uncongested flow states experience a dissipation in the non-linear case that is not present in the trapezoidal case. The delay in the clearance of cells does not appear to be a significant drawback.

Several enhancements for the discretisation of arcs were implemented in CellNetLoad. Firstly, it is possible to correctly model the arc storage capacity by increasing the storage factor of the last cell to a value between 1.0 and 2.0. This feature allows more accurate modelling of spillback. The error in the arc length is not corrected in the same fashion because it complicates the clearance of a cell. A second enhancement is the increase in the upper bound of the time step due to the generalised intersection model.

The discretisation of flow involves the specification of a vehicle unit, which may have any value larger than 10^{-5}. Packet size is restricted to integer multiples of the vehicle unit. This restriction can be enforced using either truncation or probabilistic rounding. It was found that both methods had a bias to restrict flow more than predicted by an analytical calculation of the cell-transmission model. This bias is greater for the truncation option, and it increases as the vehicle unit is increased, as the time step is reduced, and as the flow approaches the capacity. Because of this bias, CellNetLoad will not converge to the solution of the LWR hydrodynamic model with smaller discretisations. However, for large time steps, the probabilistic rounding option creates some stochasticity that seems appropriate relative to the benchmark, and the execution times are reduced because small quantities of flow are not moved each time step.

Other features of the flow propagation of the cell-transmission model which were not investigated include a variable time step, and the implementation of heuristic rules which model the capacity reduction due to lane changing. These topics are suggested for future study.

Chapter 6

Intersection Model

6.1 Introduction

The cell-transmission model is able to model diverges and merges only. Intersections with multiple arcs in the backward star *and* the forward star must be represented using a diverge on each approach, which resembles a freeway interchange. Such a representation is illustrated in Figure 4.25 for a 4-leg intersection. The generalised intersection model of CellNetLoad is an iterative algorithm which permits more realistic modelling of intersections. The basic intersection model is presented in chapter 3, and tested in chapter 4.

Two enhancements were made to the basic intersection model. Firstly, the calculation of the partial demands was modified to provide several choices in the degree of enforcement of first-in-first-out (FIFO) discipline. The second enhancement was the explicit modelling of lanes in the last cell of the arcs of the backward star to improve the modelling of shared lanes and lane blockages. These enhancements are described in this chapter and tested using the diverge network, merge network, and intersection network described in chapter 4.

6.2 The basic intersection model

The basic intersection model is an iterative algorithm which prepares the flow to depart the last cell of each arc in the backward star by placing the packets in a list, which is later moved to the downstream arc. The algorithm is presented in Figure 3.7. The discharge capacity and the local supply are distributed among the competing streams according to the proportion of the partial demand of that stream. FIFO discipline is enforced at the level of cohorts, but packets within the same cohort may overtake one another.

The intersection algorithm resolves merges and diverges as max flow problems with capacitated arcs in a graph. For example, the 4-leg intersection, which was presented in chapter 4, consists of four diverges and four merges. The corresponding flow graphs are shown in Figure 6.1. The node indices in the graph correspond to i_j for the ordinary cells receiving flow from the diverge cell with index $i \in A^-$ and sending flow to the merge cell with index $j \in A^+$.

The max flow on each path of the diverge graph corresponds to the flow y^t_{i,i_j} during time step $[t_j, t_j + \Delta t)$ of each stream ij entering the ordinary cell i_j. This flow is determined iteratively, by assigning the max flow to each path in the graph, then reducing the corresponding arc capacities of the path in the graph. The capacity of the arc from the source to node i in the diverge graph is redundant because the sum of the capacities of the arcs exiting node i is bounded from above by $\overline{Q}_i^m$ since $\sum_j \alpha_{ij}^m = 1$. The arcs with capacities defined by the min() function could be equivalently split into two arcs, each with a capacity represented by a single variable. The min() function was used for convenience of notation. The number of iterations required to determine the max flow is bounded from above by the number of paths in the diverge graph. The algorithm must converge each time the partial demands are determined. The algorithm terminates when it converges and no more partial demands are considered.

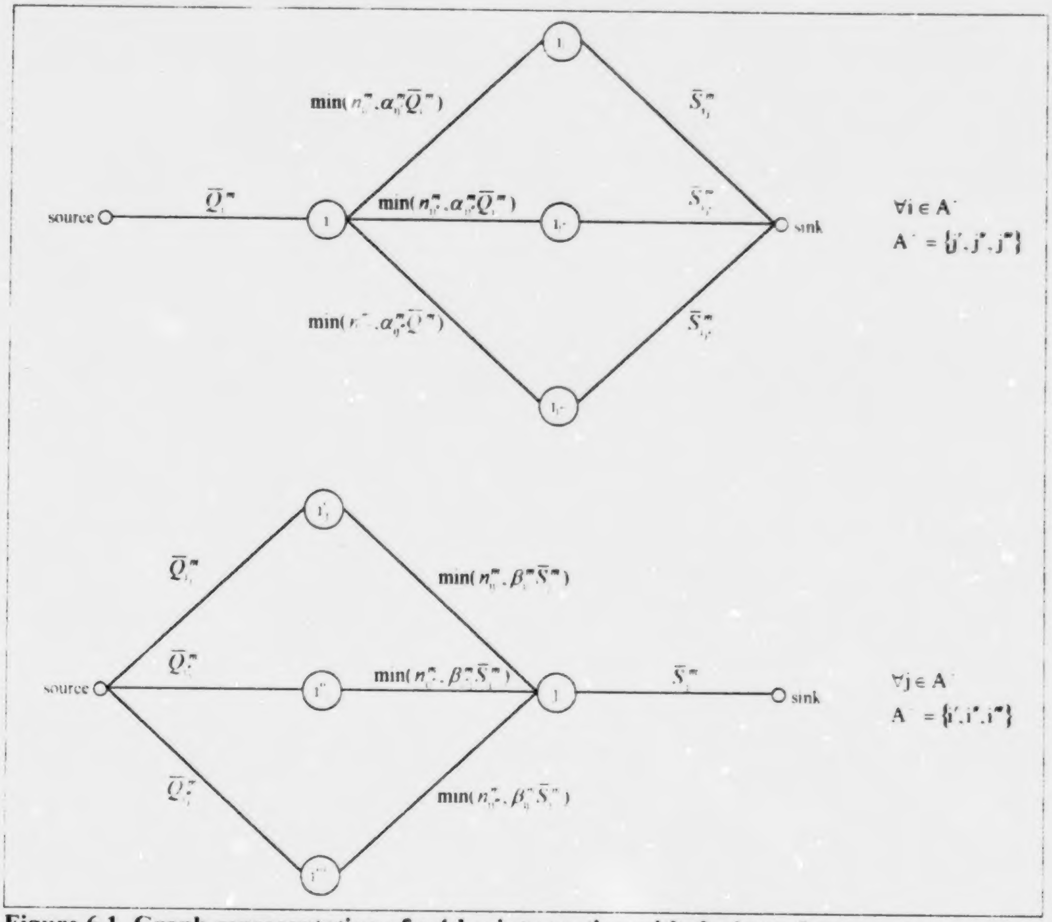

source
$\overline{Q}_i^m$
i
$\min(n_{ij'}^m, \alpha_{ij'}^m \overline{Q}_i^m)$
$\min(n_{ij''}^m, \alpha_{ij''}^m \overline{Q}_i^m)$
$\min(n_{ij'''}^m, \alpha_{ij'''}^m \overline{Q}_i^m)$
j'
j''
j'''
$\overline{S}_{j'}^m$
$\overline{S}_{j''}^m$
$\overline{S}_{j'''}^m$
sink
$\forall i \in A^+$
$A^+ = \{j', j'', j'''\}$
source
$\overline{Q}_{i'}^m$
$\overline{Q}_{i''}^m$
$\overline{Q}_{i'''}^m$
i'
i''
i'''
$\min(n_{i'j}^m, \beta_{i'j}^m \overline{S}_j^m)$
$\min(n_{i''j}^m, \beta_{i''j}^m \overline{S}_j^m)$
$\min(n_{i'''j}^m, \beta_{i'''j}^m \overline{S}_j^m)$
j
$\overline{S}_j^m$
sink
$\forall j \in A^-$
$A^- = \{i', i'', i'''\}$

Figure 6.1 Graph representation of a 4-leg intersection with the interchange representation

Proposition 6.1 The number of iterations required to determine the max flow is bounded from above by the number of paths in the diverge graph.

Proof 6.1

Flow is added to each path in the diverge graph, respecting the arc capacities.

$$\Delta y_{ij}^m = \min\left(\overline{Q}_i^m, \min(n_{ij}^m, \alpha_{ij}^m \overline{Q}_i^m), \overline{S}_j^m\right) = \min\left(n_{ij}^m, \alpha_{ij}^m \overline{Q}_i^m, \overline{S}_j^m\right) \ \forall j \in A^+$$

The second term in the min() function represents the discharge capacity shared among the streams present in the diverge cell. If the second term is an active constraint for all streams in the diverge cell (Case 1), then the discharge capacity is exhausted during this iteration.

$$\text{Case 1}: \ \forall j \in A^+ : \Delta y_{ij}^m = \alpha_{ij}^m \overline{Q}_i^m \ \ \forall j \in A^+ \Rightarrow \sum_j \Delta y_{ij}^m = \overline{Q}_i^m \Rightarrow \overline{Q}_i^{m+1} = 0$$

Alternatively, if the second term is not an active constraint for a particular stream, then one of the other terms must be an active constraint (Case 2). If the partial occupancy n_{ij}^{m+1} is an active constraint, then no vehicles from that stream will be present in the diverge cell during the next iteration. If the local supply $\overline{S}_{i_j}^{m+1}$ is an active constraint, then there will be no remaining local supply in the ordinary cell, which carries that stream, at the beginning of the next iteration.

$$\text{Case 2}: \exists \mathrm{j} \in \mathrm{A}^{+} \quad \Delta y_{ij}^{m} < \alpha_{ij}^{m} \overline{Q}_{i}^{m} \Rightarrow \begin{cases} \Delta y_{ij}^{m} = n_{ij}^{m} \Rightarrow n_{ij}^{m+1} = 0 \\ \Delta y_{ij}^{m} = \overline{S}_{i_j}^{m} \Rightarrow \overline{S}_{i_j}^{m+1} = 0 \end{cases}$$

For both Case 1 and Case 2, the number of paths with non-zero capacity is reduced by at least one during each iteration. Therefore, the number of iterations until convergence is bounded from above by the number of streams present in the diverge cell.

The same observations apply to the merge graph. The max flow on each path of the merge graph corresponds to the flow y_{ij}^{t} during time step $[t_i, t_i + \Delta t)$ of each stream ij exiting the ordinary cell i_j. The capacity of the arc from the node j to the sink is redundant because the sum of the capacities of the arcs entering node j is bounded from above by $\overline{S}_j^m$ since $\sum_i \beta_{ij}^m = 1$. The number of iterations required to determine the max flow is bounded from above by the number of paths in the merge graph.

6.2.1 Discussion of weaknesses

The evaluation in chapter 4 of the basic cell-transmission model revealed that the above model was appropriate for uncongested merges and diverges with single-lane arcs. Traffic signal phasing should be modelled explicitly. Though capacity reductions due to lane-changing and gap-acceptance at intersections cannot be modelled by the cell-transmission model, predictions from external models could be used to specify the reduced capacity in the model input. Two weaknesses of the basic cell-transmission model are the realism of modelling lane reductions and lane permissions at intersections, also called lane striping.

Lane reductions are poorly modelled in two respects. Firstly, the asymmetry of a lane drop is not captured. The asymmetry of dropping the left lane or right lane results in asymmetric lane changes, the effects of which cannot be applied to individual lanes in the cell-transmission model. Secondly, the anticipation of drivers is ignored, so that the effects take place at the location of the lane drop, instead of farther upstream where the lane changes occur. During congested conditions, most lane changes take place immediately upstream of the lane drop, but the lane changes will take place farther upstream as the congestion decreases. Because the location of the lane changes is not fixed, this phenomenon cannot be captured by altering the location of the lane drop in the model input.

The lack of sensitivity to lane striping is a more severe concern. The constraints of lane striping are relevant on multi-lane intersection approaches only if one or more streams encounters congested conditions. This congestion may result from an excess of partial demand compared to the number of available lanes for discharge, or from limited local supply of the downstream arcs. While one stream encounters congestion, other streams with access to uncongested downstream arcs via exclusive discharge lanes should not be affected immediately. Eventually, congestion may spill over across other lanes upstream of the intersection, blocking the access of other streams.

The effects of lane-striping may be modelled by explicitly separating the streams with a diverge on each approach, then specifying the share of the discharge capacity to each stream in the input files. This interchange representation of intersections was used in chapter 4. It was found that spillback cannot be modelled properly with this representation, and that the upper bound of the time step size may be reduced to one-third, thereby tripling the minimum execution time of the simulation. An alternative representation does not require the diverge on the arcs, since the iterative algorithm of CellNetLoad permits diverge cells to discharge directly into merge cells across a node. The results of this representation are presented below.

6.2.2 Alternative intersection representation

The alternative intersection representation is made possible by the generalised intersection model, which determines the flows of each stream across the intersection node during each time step. The last cell of each arc is different from the ordinary cells, which constitute the rest of the arc. The last cell may function as an ordinary cell if the backward star and the forward star contain one arc each. The last cell may also function as a diverge cell, discharging streams to each arc in the forward star. If an arc contains only a single cell, and several upstream arcs send flow to that arc, then that last cell will function as a merge cell.

The algorithm, presented in Figure 3.7, does not allow a fully parallel implementation. The flow of each stream exiting a last cell depends on the occupancy of the first cell of the arcs in the forward star, and on the partial demands of the last cell of the other arcs in the backward star. Therefore, the flow of all the last cells in the backward star of a particular node must be determined by the same procedure at the same time. This operational limitation may prove to be warranted due to improvements in accuracy and increases to the upper bound of the time step.

6.2.2.1 Graph representation and the max flow problem

Generalised intersections have a more complex graph representation, as shown in Figure 6.2. The ordinary cells operating as a buffer between the diverge cells and merge cells have been eliminated.

In the context of a max flow problem, flow may be added to any flow-augmenting path of the graph. A flow-augmenting path may be either a simple path or a complex path. A simple path begins at the source, then passes through only two nodes before ending at the sink. A simple path corresponds to a stream. A complex path contains more nodes, i.e. multiple cells from the backward star and multiple cells from the forward star. For example, adding flow to the complex path (source - i′ - j′ - i″ - j″ - sink) essentially removes an amount of flow from stream i″j′ in favour of adding the same amount of flow to streams i″j″ and i′j′.

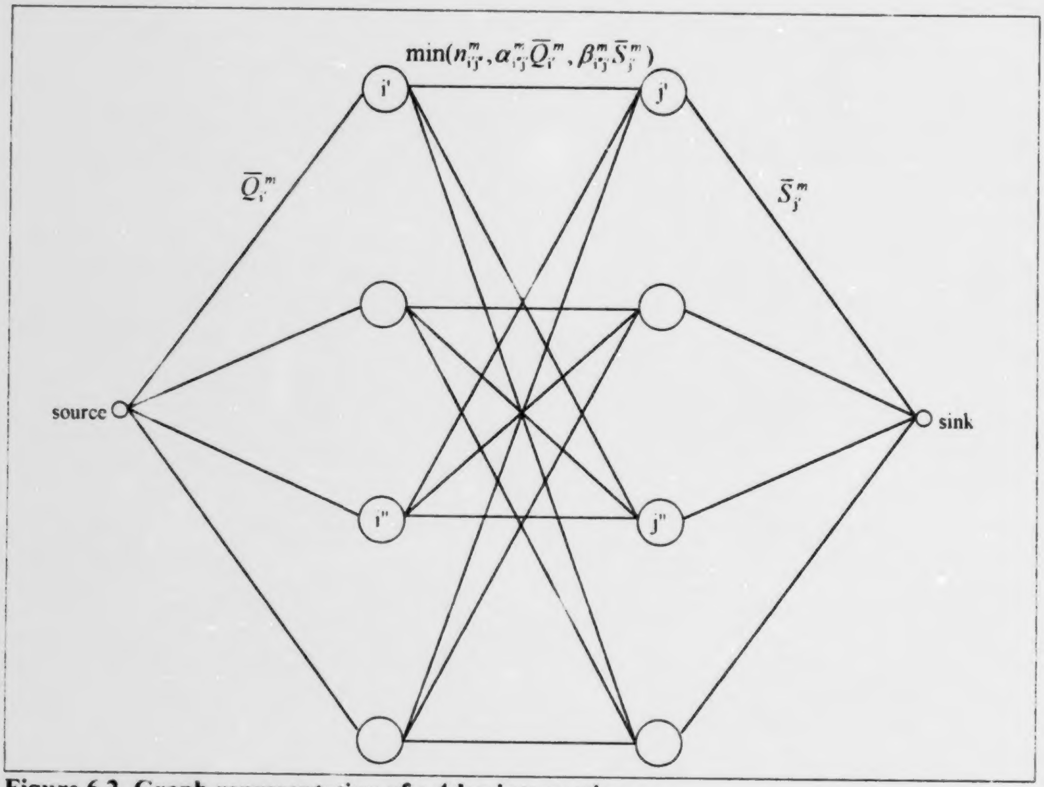

$\min(n_{i'j'}^m, \alpha_{i'j'}^m \overline{Q}_{i'}^m, \beta_{i'j'}^m \overline{S}_{j'}^m)$
i'
j'
$\overline{Q}_{i'}^m$
$\overline{S}_{j'}^m$
source
sink
i"
j"

Figure 6.2 Graph representation of a 4-leg intersection

The intersection algorithm of CellNetLoad considers each stream independently during each iteration Flow is added to the corresponding simple path in the graph, respecting the arc capacities.

$$[6.1] \qquad \Delta y_{ij}^m = \min\left(\overline{Q}_i^m, \min(n_{ij}^m, \alpha_{ij}^m \overline{Q}_i^m, \beta_{ij}^m \overline{S}_{ij}^m), \overline{S}_{ij}^m\right) \forall \mathrm{i} \in \mathrm{A}^-, \mathrm{j} \in \mathrm{A}^+.$$

Complex paths are not considered in the generalised intersection algorithm, because they are inconsistent with the selfish behaviour of drivers at traffic intersections. In real traffic networks, one stream does not sacrifice flow for the sake of other streams unless the traffic control forces such a sacrifice. Adding flow to a complex path would violate the distribution coefficients for the discharge capacity and/or the local supply. Relaxing these constraints would result in higher flows through the intersection.

An example is included in Figure 6.3. A single iteration of the generalised intersection algorithm yields a solution with total flow of 20 vehicles. The capacity constraints implied by the distribution coefficients are then relaxed, and the max flow solution yields a total flow of 25 vehicles, but requires stream $i''j'$ to sacrifice flow for streams $i''j''$ and $i'j'$.

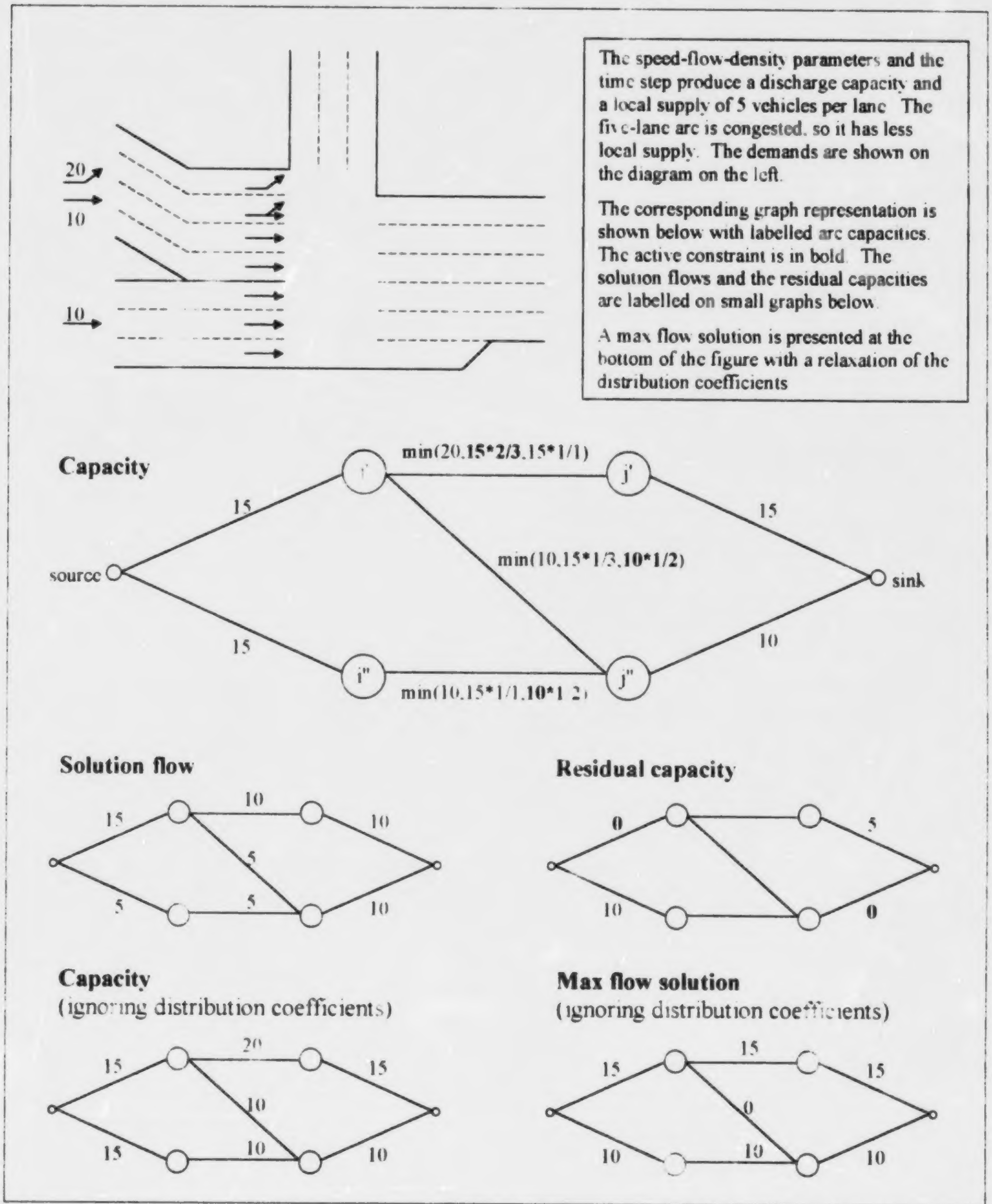
6.2.2.2 Convergence of the generalised intersection algorithm

This section extends Proof 6.1 for generalised intersections with the alternative representation. Proof 6.1 is valid for intersections with graph representations that consist of simple paths only. Intersections with multiple arcs in the backward star *and* multiple arcs in the forward star have graph representations that may include complex paths. The graph of a 4-leg intersection with the alternative intersection representation is shown in Figure 6.2.

The intersection algorithm adds flow to each simple paths, then updates the arc capacities during each iteration. If at least one arc in each simple path of the graph has a capacity equal to zero, then the algorithm has converged. This can be stated mathematically as

$$[6.2] \qquad \forall i \in A^-, j \in A^+ : n_{ij}^m = 0 \text{ or } \sum_j \Delta y_{ij}^m = \overline{Q}_i^m \text{ or } \sum_j \Delta y_{ij}^m = \overline{S}_{ij}^m .$$

After convergence, the algorithm checks if the front cohort has been depleted on each arc in the backward star. If so, the next cohort in the cohort queue (assuming one exists) is placed at the front of the cohort queue, and a new set of partial demands are derived from it. The algorithm terminates when the updated partial demands remain the same. A proof by contradiction using [6.1] and [6.2] can show that the number of iterations of the intersection algorithm is finite.



The speed-flow-density parameters and the time step produce a discharge capacity and a local supply of 5 vehicles per lane The five-lane arc is congested, so it has less local supply. The demands are shown on the diagram on the left.

The corresponding graph representation is shown below with labelled arc capacities. The active constraint is in bold. The solution flows and the residual capacities are labelled on small graphs below.

A max flow solution is presented at the bottom of the figure with a relaxation of the distribution coefficients

20
10
10
Capacity
min(20,15*2/3,15*1/1)
i'
j'
15
15
source
min(10,15*1/3,10*1/2)
sink
15
10
i''
j''
min(10,15*1/1,10*1/2)
Solution flow
10
15
10
5
5
5
10
Residual capacity
0
5
10
0
Capacity
(ignoring distribution coefficients)
20
15
15
10
15
10
10
Max flow solution
(ignoring distribution coefficients)
15
15
15
0
10
10
10

Figure 6.3 Example of an intersection for which the generalised intersection algorithm does not find the max flow solution when the distribution of the discharge capacity and local supply are relaxed

Proposition 6.2 The number of iterations required for convergence is bounded from above by the number of streams at the intersection.

Proof 6.2

Consider the first iteration of the generalised intersection algorithm for an intersection with multiple arcs in the backward star and multiple arcs in the forward star. All arcs in the graph have positive capacities. (Any arcs with zero capacity can be safely removed from the graph without loss of generality.) It is assumed that some complex path(s) is present. If not, then a proof similar to Proof 6.1 would be sufficient.

Hypothesis 1: After one iteration, no arcs in the graph have zero capacity, i.e.

$$\Delta y_{ij}^m < n_{ij}^m \qquad \forall i \in A^-, j \in A^+ \qquad \text{[H1a]}$$

$$\sum_j \Delta y_{ij}^m < \overline{Q}_i^m \qquad \forall i \in A^- \qquad \text{[H1b]}$$

$$\sum_i \Delta y_{ij}^m < \overline{S}_j^m \qquad \forall j \in A^+ \qquad \text{[H1c]}$$

From [6.1], $\Delta y_{ij}^m = \min(n_{ij}^m, \alpha_{ij}^m \overline{Q}_i^m, \beta_{ij}^m \overline{S}_{ij}^m)\ \forall i \in A^-, j \in A^+$

$$\text{[H1a]} \Rightarrow \Delta y_{ij}^m = \min(\alpha_{ij}^m \overline{Q}_i^m, \beta_{ij}^m \overline{S}_{ij}^m)\ \forall i \in A^-, j \in A^+ \qquad \text{[H1a']}$$

Now, assume that $\exists i \in A^- : \Delta y_{ij}^m = \alpha_{ij}^m \overline{Q}_i^m \ \ \forall j \in A^+$, which implies that $\sum_j \Delta y_{ij}^m = \overline{Q}_i^m \sum_j \alpha_{ij}^m = \overline{Q}_i^m$. This contradicts [H1b]. So, the assumption is false, and the following property must be true

$$\forall i \in A^- : \exists j \in A^+ : \Delta y_{ij}^m < \alpha_{ij}^m \overline{Q}_i^m. \qquad \text{[P1]}$$

Similarly, the assumption $\exists j \in A^+ : \Delta y_{ij}^m = \beta_{ij}^m \overline{S}_j^m \ \ \forall i \in A^-$ is false, so

$$\forall j \in A^+ : \exists i \in A^- : \Delta y_{ij}^m < \beta_{ij}^m \overline{S}_j^m. \qquad \text{[P2]}$$

From [H1a'], properties [P1] and [P2] imply that

$$\forall i \in A^- : \exists j \in A^+ : \alpha_{ij}^m \overline{Q}_i^m > \beta_{ij}^m \overline{S}_j^m = \Delta y_{ij}^m \qquad \text{[P1']}$$

$$\forall j \in A^+ : \exists i \in A^- : \Delta y_{ij}^m = \alpha_{ij}^m \overline{Q}_i^m < \beta_{ij}^m \overline{S}_j^m \qquad \text{[P2']}$$

Now, choose any $i' \in A^-$.

$$\text{[P1']} \Rightarrow \exists j' \in A^+ : \alpha_{i'j'}^m \overline{Q}_{i'}^m > \beta_{i'j'}^m \overline{S}_{j'}^m \Rightarrow \frac{n_{i'j'}^m}{\sum_j n_{i'j}^m} \overline{Q}_{i'}^m > \frac{n_{i'j'}^m}{\sum_i n_{ij'}^m} \overline{S}_{j'}^m \Rightarrow \frac{\overline{Q}_{i'}^m}{\sum_j n_{i'j}^m} > \frac{\overline{S}_{j'}^m}{\sum_i n_{ij'}^m}$$

$$\text{[P2']} \Rightarrow \exists i'' \in A^- : \alpha_{i''j'}^m \overline{Q}_{i''}^m < \beta_{i''j'}^m \overline{S}_{j'}^m \Rightarrow \frac{\overline{Q}_{i''}^m}{\sum_j n_{i''j}^m} < \frac{\overline{S}_{j'}^m}{\sum_i n_{ij'}^m}$$

If $i'' = i'$, then a contradiction is evident. So, $i'' \neq i'$, and combining the above yields

$$\frac{\overline{Q}_{i'}^m}{\sum_j n_{i'j}^m} < \frac{\overline{Q}_{i''}^m}{\sum_j n_{i''j}^m}.$$

Similarly,

$$[\mathrm{P1'}] \Rightarrow \exists j'' \in A^{+}, j'' \neq j' : \alpha^{m}_{i'j''}\overline{Q}^{m}_{i'} > \beta^{m}_{i'j''}\overline{S}^{m}_{j''} \Rightarrow \frac{\overline{Q}^{m}_{i'}}{\sum_{j} n^{m}_{i'j}} > \frac{\overline{S}^{m}_{j''}}{\sum_{i} n^{m}_{ij''}}$$

$$[\mathrm{P2'}] \Rightarrow \exists i''' \in A^{-}, i''' \neq i'' : \alpha^{m}_{i'''j''}\overline{Q}^{m}_{i'''} < \beta^{m}_{i'''j''}\overline{S}^{m}_{j''} \Rightarrow \frac{\overline{Q}^{m}_{i'''}}{\sum_{j} n^{m}_{i'''j}} < \frac{\overline{S}^{m}_{j''}}{\sum_{i} n^{m}_{ij''}}$$

These can be combined again to yield $\frac{\overline{Q}^{m}_{i'}}{\sum_{j} n^{m}_{i'j}} < \frac{\overline{Q}^{m}_{i''}}{\sum_{j} n^{m}_{i''j}} < \frac{\overline{Q}^{m}_{i'''}}{\sum_{j} n^{m}_{i'''j}}$.

Obviously, if $i''' = i''$, then a contradiction exist. So, $i''' \neq i'' \neq i'$.

This process can be repeated indefinitely, because [P1'] and [P2'] are true for all $i \in A^{-}$ and $j \in A^{+}$. But, the number of arcs in the backward star and the forward star is finite, so a contradiction is inevitable. Thus, Hypothesis 1 must be false.

The negation of Hypothesis 1 is true, so

$$\exists i' \in A^{-}, j' \in A^{+} : \Delta v^{m}_{i'j'} = n^{m}_{i'j'} \text{ or } \sum_{j} \Delta v^{m}_{i'j} = \alpha^{m}_{i'j'}\overline{Q}^{m}_{i'} \text{ or } \sum_{i} \Delta v^{m}_{ij'} = \beta^{m}_{i'j'}\overline{S}^{m}_{j'}.$$

After updating the partial demand, the remaining discharge capacity, and the remaining local supply, this path will contain an arc with zero capacity. This arc may be removed from the graph for subsequent iterations. After $\left\|A^{-}\right\| \times \left\|A^{+}\right\|$ iterations, the graph will contain no simple paths, so the algorithm will have converged.

Since the number of cohorts in the last cell of the arcs in the backward star is finite, the algorithm terminates in a finite number of iterations. Consequently, the total number of iterations until termination of the generalised intersection algorithm is bounded from above by $\left\|A^{-}\right\| \times \left\|A^{+}\right\| \times \sum_{i \in A^{-}} n_{i}^{\text{cohorts}}$, where n_{i}^{cohorts} is the number of cohorts in the cohort queue of the last cell i in the backward star.

The generalised intersection algorithm used in CellNetLoad is a greedy algorithm that converges to a max flow solution in a finite number of iterations, according to Proposition 6.2. Proof 6.2 relies on the specification of the distribution coefficients in proportion to the partial demands. If another distribution is specified, then Proof 6.2 may not apply. In such a case, the number of iterations required for convergence may be infinite.

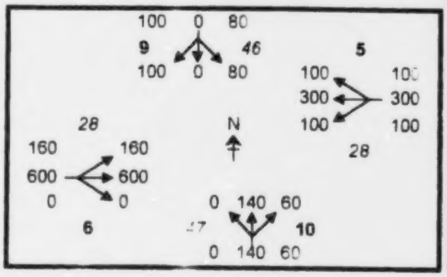
6.2.2.3 Simulation results for the alternative intersection representation

The simulation results of the intersection network, presented in chapter 4, are shown in Figure 6.4 for the alternative representation described above. Signal phasing was modelled explicitly, no capacity reductions were specified, and the time step was specified as 20 s. There was no oversaturation, the average travel time included only the travel time at free speed and the delay due to cyclic queuing. When the demand was doubled from light flow to heavy flow, the density more than doubled. The incident had no effect, despite moving its location 100 m upstream to account for the drivers' anticipation when changing lanes to avoid the incident. Reducing the time step to 5 s led to a change in some flows and densities in the second decimal place, and to some travel times in the first decimal place.

With this alternative representation, reduced capacities could only be specified for the arcs in the backward star, and not to individual streams. The reduced capacity was calculated by summing the average arc outflow, then dividing it by the number of lanes and the ratio of the length of the green interval to the cycle length.

The simulation results for intersection with specified capacity reductions are shown in Figure 6.5. The light flow scenario shows slight increases to the density and travel time due to the less intense discharge of the platoon. In the heavy flow scenario, all movements are oversaturated except for the northbound approach. With the incident, the eastbound approach is no longer oversaturated, and the westbound approach is more oversaturated. Note that all movements on an oversaturated approach are equally oversaturated. The left turn movement cannot be penalised more than the other streams. The queuing is increased on the oversaturated approaches, leading to higher densities and travel times.

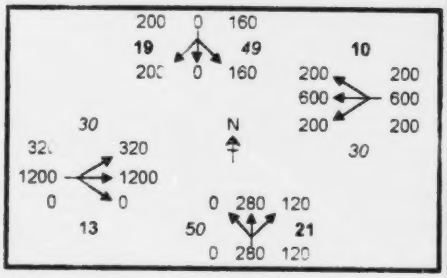
Light Flow


100 0 80
9 46
100 0 80
5
100 300 100
100 300 100
28
N
160 600 0
160 600 0
28
0 140 60
6 47 10
0 140 60

approach	demand (vph)		
	left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	160	600	0	6.3	28
W bound	100	300	100	4.9	28
S bound	80	0	100	8.7	46
N bound	0	140	60	9.7	47

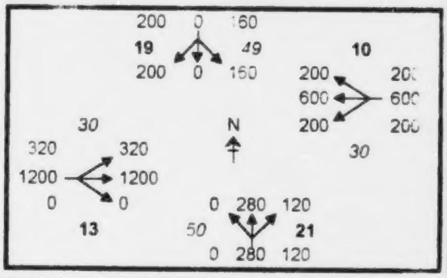
Heavy flow


200 0 160
19 49
200 0 160
10
200 600 200
200 600 200
30
N
320 1200 0
320 1200 0
30
0 280 120
13 50 21
0 280 120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	13.3	30
W bound	200	600	200	10.2	30
S bound	160	0	200	18.5	49
N bound	0	280	120	20.9	50

Heavy flow with an incident


200 0 160
19 49
200 0 160
10
200 600 200
200 600 200
30
N
320 1200 0
320 1200 0
30
0 280 120
13 50 21
0 280 120

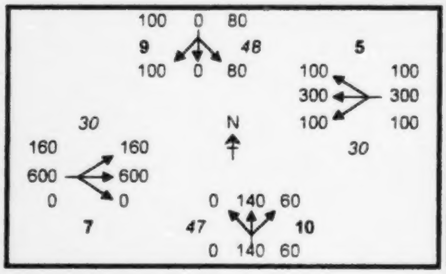
approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	13.3	30
W bound	200	600	200	10.2	30
S bound	160	0	200	18.5	49
N bound	0	280	120	20.9	50

Figure 6.4 Simulation results from CellNetLoad of an intersection

The effect of the incident was to reduce the density and travel time on the eastbound approach because the backward shockwave never reached the intersection despite moving the incident 100 m upstream. In fact, the reduced capacity actually decreases the speed of the backward shockwave because the density upstream of the incident is less than the density at capacity.

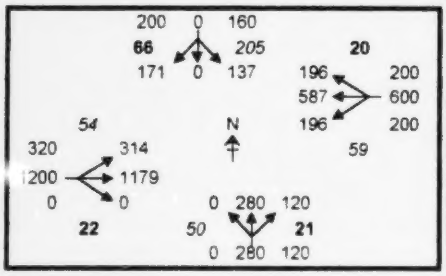
Light Flow


100
0
80
9
48
5
100
0
80
100
100
300
300
30
N
100
100
160
160
30
600
600
0
0
0
140
60
7
47
10
0
140
60

approach	demand (vph)		
	left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	160	600	0	**6.6**	*30*
W bound	100	300	100	**5.2**	*30*
S bound	80	0	100	**9.0**	*48*
N bound	0	140	60	**9.7**	*47*

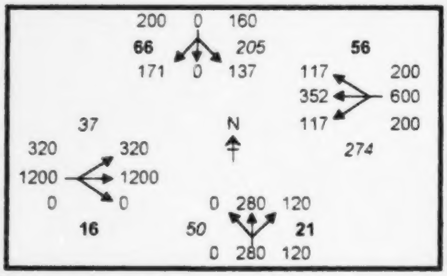
Heavy flow


200
0
160
66
205
20
171
0
137
196
200
587
600
54
N
196
200
320
314
59
1200
1179
0
0
0
280
120
22
50
21
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	314	1179	0	**21.6**	*54*
W bound	196	587	196	**20.1**	*59*
S bound	137	0	171	**65.9**	*205*
N bound	0	280	120	**20.9**	*50*

Heavy flow with an incident


200
0
160
66
205
56
171
0
137
117
200
352
600
37
N
117
200
320
320
274
1200
1200
0
0
0
280
120
16
50
21
0
280
120

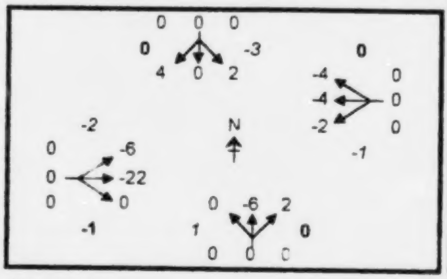
approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	**15.8**	*37*
W bound	117	352	117	**55.8**	*274*
S bound	137	0	171	**65.9**	*205*
N bound	0	280	120	**20.9**	*50*

Figure 6.5 Simulation results from CellNetLoad of an intersection with reduced capacity

The simulation results of INTEGRATION were subtracted from those of CellNetLoad for comparison to the benchmark in Figure 6.6. The left turn flows tend to be overestimated at the expense of the other movements under heavy flow, as a result of the arc capacity reduction. The density and travel time estimates for the heavy flow scenario match those from INTEGRATION relatively well, though these measures are still underestimated on the eastbound approach even though the arc outflow matched that of INTEGRATION. The same is true for the eastbound and westbound approaches in the incident scenario.

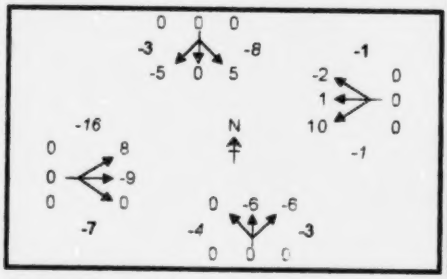
Light Flow



approach	demand (vph)		
	left	through	right
E bound	0	0	0
Wbound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	-6	-22	0	-0.7	-2
Wbound	-2	-4	-4	-0.2	-1
S bound	2	0	4	-0.4	-3
N bound	0	-6	2	-0.2	1

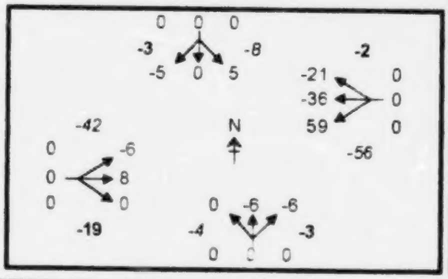
Heavy flow



approach	demand (vph)		
	left	through	right
E bound	0	0	0
Wbound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	8	-9	0	-6.6	-16
Wbound	10	1	-2	-0.6	-1
S bound	5	0	-5	-3.4	-8
N bound	0	-6	-6	-3.0	-4

Heavy flow with an incident



approach	demand (vph)		
	left	through	right
E bound	0	0	0
Wbound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	-6	8	0	-18.8	-42
Wbound	59	-36	-21	-2.4	-56
S bound	5	0	-5	-3.4	-8
N bound	0	-6	-6	-3.0	-4

Figure 6.6 Simulation results from CellNetLoad of an intersection with reduced capacity compared to INTEGRATION

6.3 Relaxation of cohort FIFO discipline

The first enhancement to the intersection model involves the relaxation of the FIFO condition. The FIFO condition must be enforced strictly in flow-based models to prevent overtaking within the same traffic stream. Unfortunately, this strict respect of FIFO does not allow overtaking among streams which share the same arc. This is unrealistic for multi-lane intersection approaches, when congestion builds on one stream and other streams have exclusive permission on other lanes. If FIFO is strictly enforced under these circumstances, then it is as if the model predicts instantaneous spillover of any queue across all lanes of the approach.

In CellNetLoad, the default intersection logic enforced cohort FIFO discipline. Overtaking of one packet over another is permitted within a cohort, but no overtaking takes place between different cohorts. This is the most restrictive enforcement of FIFO discipline in CellNetLoad.

6.3.1 Implementation

The prototype code offers a choice of two more levels of greater relaxation of FIFO discipline by altering the value of a flag on the command line. The first option is the enforcement of stream FIFO discipline. Packets within the last cell of each arc may overtake other packets in the same cell, but only if the other packet is in another stream. The implementation of stream FIFO required alterations to step 1 and step 4 of the algorithm in Figure 3.7.

The new algorithm is presented in Figure 6.7. Step 1 is simplified by adding the occupancies of each stream in the last cell i, then multiplying by the ratio of the local demand to the total occupancy only if the local demand is exceeded. Packets are not considered for partial demand if the downstream remaining local supply is zero. No cohorts are split during this step of the algorithm. Step 4 is complicated by the removal of packets from different cohorts during each iteration. The first case deals with cohorts closest to the front of the queue. All packets going to downstream cell j from these cohorts are removed. In the second case, the flow is satisfied by sending only a portion of each packet from this cohort to downstream cell j, so only the remainder of the packets is available for the next iteration. In the third case, no packets are removed because the cohorts are far enough back in the cohort queue.

Figure 6.7 Cell FIFO algorithm for the determination of intersection flow – overtaking permitted among packets in the same cell

Step 0. Initialisation

iteration $m = 0$

vehicle movement $y_{ij}^{m} = 0 \quad \forall i \in A^-, j \in A^+$

local demand remainder $\overline{D}_i^{m} = D_i \quad \forall i \in A^-$

discharge capacity remainder $\overline{Q}_i^{m} = Q_i \quad \forall i \in A^-$

local supply remainder $\overline{S}_j^{m} = S_j \quad \forall j \in A^+$

packet groups to arcs in forward star $n_{ij\tau}^{m} = \sum_d n_{ij\tau}^{d} \quad \forall i \in A^-, j \in A^+, \tau \in (0, \bar{T})$

Step 1. Calculate partial demand

$$n_{ij}^{m} = \begin{cases} 0 & \text{if } \overline{S}_j^{m} = 0 \\ \sum_\tau n_{ij\tau}^{m} & \text{if } \overline{S}_j^{m} > 0 \text{ and} \sum \sum_\tau n_{ij\tau}^{m} \le \overline{D}_i^{m} \\ \sum_\tau n_{ij\tau}^{m} \left(\overline{D}_i^{m} \Big/ \sum_j \sum_\tau n_{ij\tau}^{m} \right) & \text{otherwise} \end{cases} \quad \forall i \in A^-, \; j \in A^+$$

Step 2. Calculate the number of vehicles Δy_{ij}^{m} to depart cell i for cell j

$$\Delta y_{ij}^{m} = \min\left(n_{ij}^{m}, \alpha_{ij}^{m} \overline{Q}_i^{m}, \beta_{ij}^{m} \overline{S}_j^{m} \right) \quad \forall i \in A^-, j \in A^+$$

where $\alpha_{ij}^{m} = n_{ij}^{m} \Big/ \sum_j n_{ij}^{m}$, $\beta_{ij}^{m} = n_{ij}^{m} \Big/ \sum_i n_{ij}^{m} \quad \forall i \in A^-, j \in A^+$

Step 3. Stopping criterion

If $\exists i \in A^-, j \in A^+ : \Delta y_{ij}^{m} > 0$, then continue to Step 4. Otherwise, stop.

Step 4. Update

$$y_{ij}^{m+1} = y_{ij}^{m} + \Delta y_{ij}^{m} \quad \forall i \in A^-, j \in A^+$$

$$\overline{D}_i^{m+1} = D_i - \sum_j y_{ij}^{m+1}, \; \overline{Q}_i^{m+1} = Q_i - \sum_j y_{ij}^{m+1} \quad \forall i \in A^-$$

$$\overline{S}_j^{m+1} = S_j - \sum_i y_{ij}^{m+1} \quad \forall j \in A^+$$

$$n_{ij\tau}^{m+1} = \begin{cases} 0 & \text{if } \sum_{\tau' \le \tau} n_{ij\tau'}^{m} \le \Delta y_{ij}^{m} \\ \sum_{\tau' \le \tau} n_{ij\tau'}^{m} - \Delta y_{ij}^{m} & \text{if } \sum_{\tau' \le \tau} n_{ij\tau'}^{m} > \Delta y_{ij}^{m} > \sum_{\tau' < \tau} n_{ij\tau'}^{m} \\ n_{ij\tau}^{m} & \text{if } \quad \Delta y_{ij}^{m} \le \sum_{\tau' < \tau} n_{ij\tau'}^{m} \end{cases} \quad \forall i \in A^-, \; j \in A^+, \; \tau \in (0, \bar{T})$$

$m \leftarrow m + 1$

Return to Step 1.

The algorithm in Figure 6.7 is equivalent to assuming that a packet in one stream is always able to overtake packets from another stream, and that no congested stream may spill over to block that overtaking, no matter how long the queue. This model is appropriate for multi-lane approaches in which each stream has exclusive permission to at least one lane. This algorithm has two drawbacks. Firstly, this model is sensitive to the cell length, since overtaking occurs only in the last cell of each arc. The amount of overtaking is reduced as the simulation time step—and hence the cell length and cell occupancy—is reduced.

Proposition 6.2 is also valid for the cell FIFO algorithm. The total number of iterations is bounded from above by $\|A^-\| \times \|A^+\|$, since the partial demands are non-increasing. The initial determination of the partial demands requires more execution time since a pass of the entire cohort queue of each last cell in the backward star must be performed.

The alternative option possible in this prototype code is to relax FIFO only at the front of the cohort queue. The local demand remainder is known at the beginning of each iteration. Thus, the partial demands are evaluated by scanning only enough cohorts from the front of the cohort queue until the local demand is satisfied. This is equivalent to assuming that overtaking may occur only among cohorts which may reach the stop line during uncongested conditions, and that spillover can block those cohorts further back in the cohort queue. This model is also sensitive to the size of the time step, but the efficiency is greatly improved.

The implementation of this alternative model required alterations to step 0 and step 1 of Figure 6.7. As in the algorithm of Figure 3.7 a cohort is split to separate the local demand from the rest of queue prior to the determination of the partial demands. These alterations are presented in Figure 6.8.

Figure 6.8 Relaxed cohort FIFO algorithm for the determination of intersection flow – overtaking permitted among the packets at the front of the cohort queue which represent the local demand

```
Step 0. Initialisation
    iteration $m = 0$
    vehicle movement $y_{ij}^{m} = 0 \ \forall i \in A^{-}, j \in A^{+}$
    local demand remainder $\overline{D}_i^{m} = D_i \ \forall i \in A^{-}$
    discharge capacity remainder $\overline{Q}_i^{m} = Q_i \ \forall i \in A^{-}$
    local supply remainder $\overline{S}_j^{m} = S_j \ \forall j \in A^{+}$
    packet groups to arcs in forward star $n_{ij\tau}^{m} = \sum_d n_{ij\tau}^{d} \ \forall i \in A^{-}, j \in A^{+}, \tau \in (0, \bar{T})$
    Check that the local demand can be separated from the cohort queue:
    $\forall i \in A^{-}$ : if $\sum_{\tau \le \bar{\tau}_i} \sum_j n_{ij\tau}^{m} > \overline{D}_i^{m}$ then split the cohort with time stamp $\bar{\tau}_i$
        into two cohorts with time stamps $\bar{\tau}_{i1} = \bar{\tau}_i$ and $\bar{\tau}_{i2} = \bar{\tau}_i + \varepsilon$
        such that $\sum_{\tau \le \bar{\tau}_{i1}} \sum_j n_{ij\tau}^{m} = \overline{D}_i^{m}$
    where $\bar{\tau}_i$ is the time stamp of the cohort of cell i which must be split
Step 1. Calculate the partial demand
    $n_{ij}^{m} = \begin{cases} 0 & \text{if } \overline{S}_j^{m} = 0 \\ \sum_{\tau \le \bar{\tau}} n_{ij\tau}^{m} & \text{if } \overline{S}_j^{m} > 0 \end{cases} \ \forall i \in A^{-}, j \in A^{+}$
Step 2. Calculate the number of vehicles $\Delta y_{ij}^{m}$ to depart cell i for cell j
    $\Delta y_{ij}^{m} = \min\left(n_{ij}^{m}, \alpha_{ij}^{m} \overline{Q}_i^{m}, \beta_{ij}^{m} \overline{S}_j^{m}\right) \ \forall i \in A^{-}, j \in A^{+}$
    where $\alpha_{ij}^{m} = n_{ij}^{m} / \sum_j n_{ij}^{m}$, $\beta_{ij}^{m} = n_{ij}^{m} / \sum_i n_{ij}^{m} \ \forall i \in A^{-}, j \in A^{+}$
Step 3. Stopping criterion
    If $\exists i \in A^{-}, j \in A^{+} : \Delta y_{ij}^{m} > 0$, then continue to Step 4. Otherwise, stop.
Step 4. Update
    $y_{ij}^{m+1} = y_{ij}^{m} + \Delta y_{ij}^{m} \ \forall i \in A^{-}, j \in A^{+}$
    $\overline{D}_i^{m+1} = D_i - \sum_j y_{ij}^{m+1}$, $\overline{Q}_i^{m+1} = Q_i - \sum_j y_{ij}^{m+1} \ \forall i \in A^{-}$
    $\overline{S}_j^{m+1} = S_j - \sum_i y_{ij}^{m+1} \ \forall j \in A^{+}$
```

$$n_{ij\tau}^{m+1} = \begin{cases} 0 & \text{if } \sum_{\tau' \le \tau} n_{ij\tau'}^{m} \le \Delta y_{ij}^{m} \\ \sum_{\tau' \le \tau} n_{ij\tau'}^{m} - \Delta y_{ij}^{m} & \text{if } \sum_{\tau' \le \tau} n_{ij\tau'}^{m} > \Delta y_{ij}^{m} > \sum_{\tau' < \tau} n_{ij\tau'}^{m} \\ n_{ij\tau}^{m} & \text{if } \qquad \Delta y_{ij}^{m} \le \sum_{\tau' < \tau} n_{ij\tau'}^{m} \end{cases} \left. \vphantom{\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}} \right\} \begin{matrix} \forall i \in A^{-}, \\ j \in A^{+}, \\ \tau \in (0, \bar{t}) \end{matrix}$$

$m \leftarrow m + 1$

Return to Step 1.

The number of iterations until termination of the relaxed cohort FIFO algorithm is bounded from above by $\|A^{-}\| \times \|A^{+}\|$, as in the cell FIFO algorithm. However, the initial determination of the partial demand is shorter than the cell FIFO algorithm during congested conditions, because only enough cohorts from each cohort queue are considered to satisfy the local demand.

6.3.2 *Effects on the diverge network*

The enforcement of cohort FIFO discipline on the merge network resulted in the unstable distribution of the local demand. The instability is manifested in the outflow from the diverge during the oversaturation demand, which is not smooth in Figure 4.18. This instability was caused by the varying size of the cohort at the front of the cohort queue on the arc upstream of the diverge.

The relaxation of cohort FIFO discipline smoothed the outflow because the partial demands were derived only once from the aggregated cohorts, instead of repeated derivations from individual cohorts in the cohort queue. The simulation results for CellNetLoad with relaxed FIFO discipline are shown in Figure 6.9. A queue of right turners grew in the last cell of the from 1240 to 1540 s in the simulation with the 20-s time step. The queue was not large enough to restrict the inflow. The left turners and through vehicles were permitted to overtake the queue of right turners.

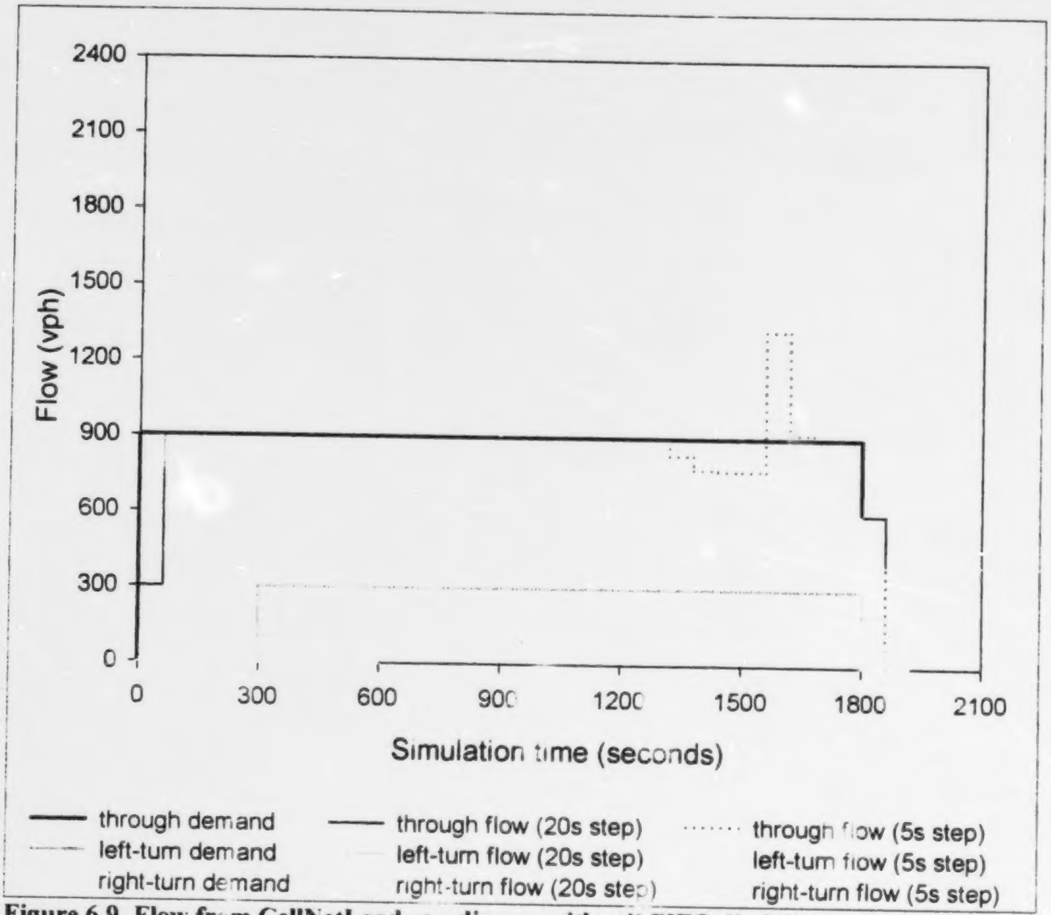


Figure 6.9 Flow from CellNetLoad at a diverge with cell FIFO discipline

In the simulation with a 5-s time step, the inflow to the diverge cell was restricted due to the queue of right turners. Since FIFO discipline is enforced between cells, some left turners and through vehicles were unable to overtake the queue of right turners, and the outflow from the diverge did not match the demand for any stream. The decreased outflow was stable, unlike Figure 4.18.

If cohort FIFO discipline was relaxed only among the cohorts at the front of the cohort queue, which represented the local demand, then the left turners and through vehicles were restricted as soon as the cohort queue exceeded the local demand. This restriction began earlier for the simulation with a 5-s time step compared to the simulation with a 20-s time step, as shown in

Figure 6.10. Again, the instability, which was found when cohort FIFO discipline was enforced strictly, due to the dependence of the partial demand on the size and composition of the front cohort only was not present for this case of relaxed FIFO discipline.

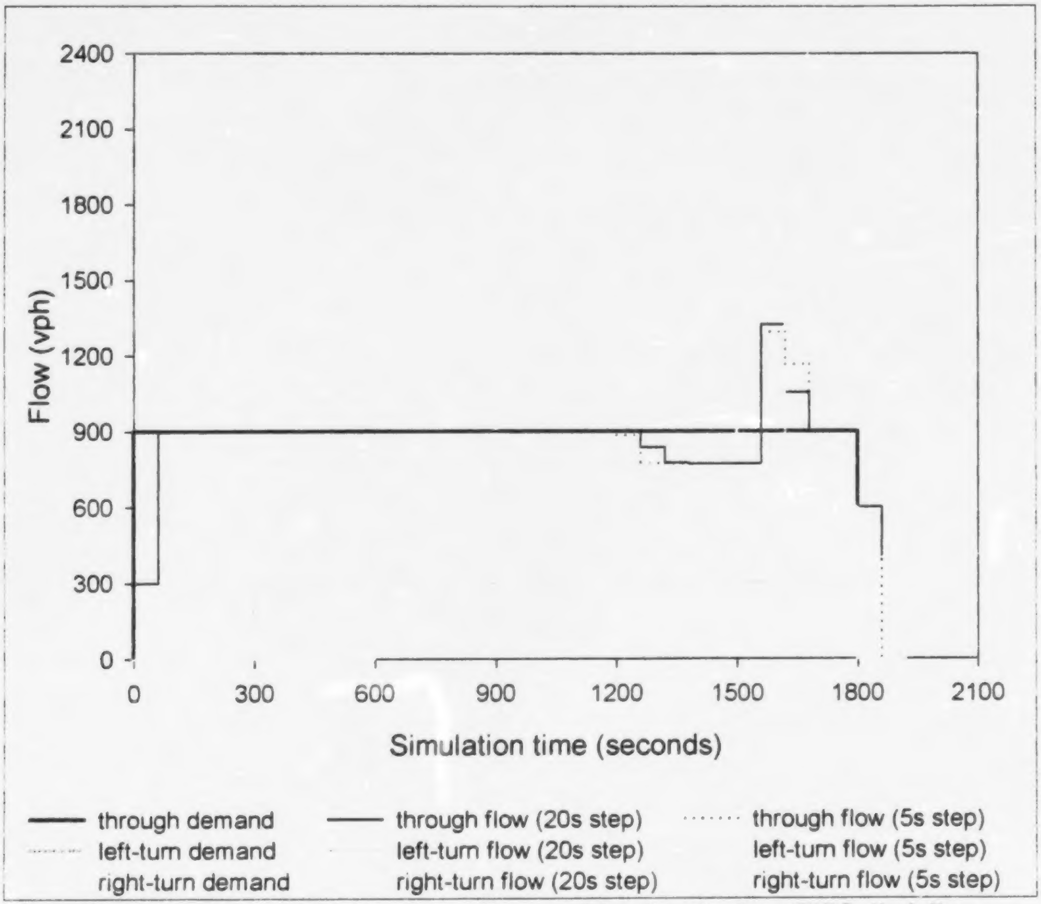


Figure 6.10 Flow from CellNetLoad at a diverge with relaxed cohort FIFO discipline

A qualitative comparison with the INTEGRATION simulation results in Figure 4.17 reveals that CellNetLoad is unrealistically efficient in the distribution of the discharge capacity. The right lane behaves like an exclusive right-turn lane immediately as the saturation demand arrives at the diverge, whereas some through vehicles chose to use the right lane in INTEGRATION. In addition, CellNetLoad underestimates the capacity reduction due to spillover relative to the benchmark.

6.3.3 *Effects on the merge network*

The enforcement of cohort FIFO discipline on the merge network resulted in the unstable distribution of the local supply. This was evident in Figure 4.22 for the observations of the flows into the merge during the simulation with the 20-s time step. The instability was caused by the dependency of the partial demands on the front cohort, and the arrival of split cohorts at the front of the cohort queue. This instability was eliminated for cell FIFO discipline since cohorts were no longer split to enforce cohort FIFO discipline. The observed flows are shown in Figure 6.11. As the queues filled the upstream cells, the inflows converged to 900 vph more smoothly than in Figure 4.22.

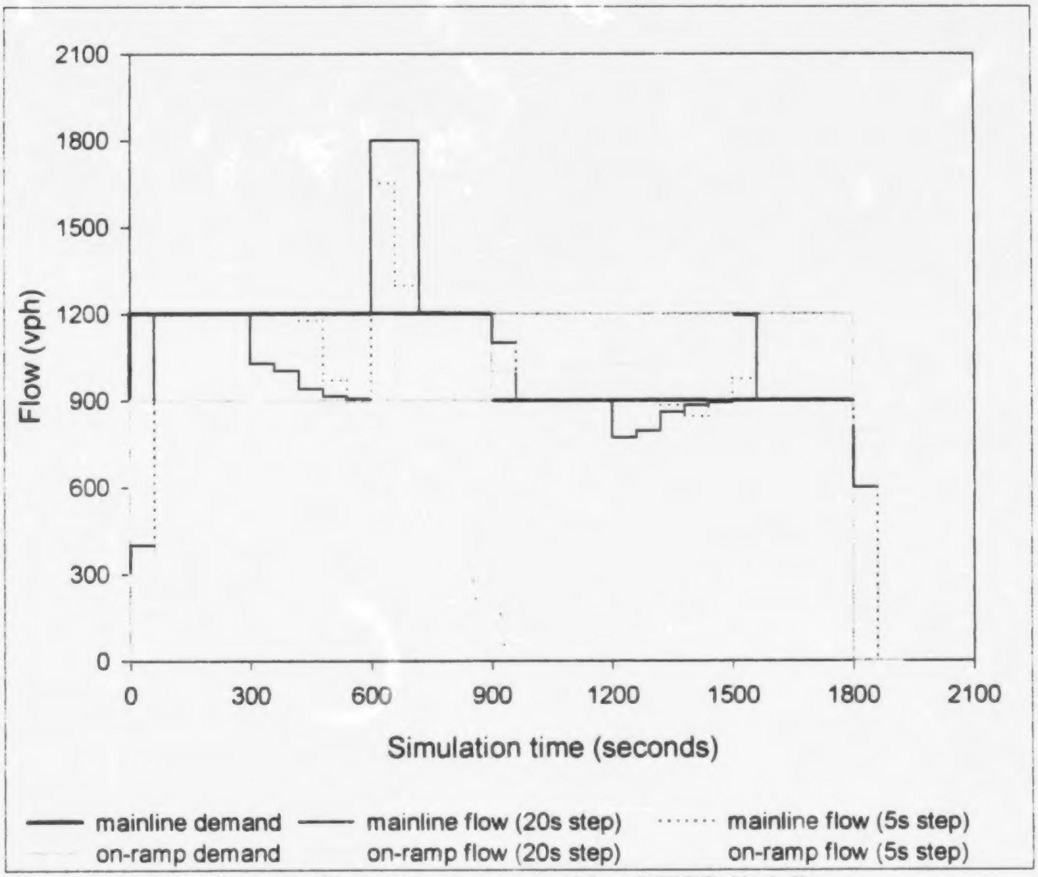


Figure 6.11 Flow from CellNetLoad at a merge with cell FIFO discipline

The alternative relaxation of cohort FIFO discipline permits overtaking only among the cohorts at the front of the cohort queue which represent the local demand. A cohort may be split so that only the packets within the local demand are considered as candidates for the partial demand. But the instability is not present because the partial demand is aggregated for all packets within the local demand, and not just for the front cohort. There was no difference between the simulation results of the two alternative relaxations.

The relaxation of FIFO discipline had less impact on the observations from the simulations with a 5-s time step, which was expected because overtaking is more restricted for smaller cells. The queue reached the first cell on the arc downstream of the merge almost three minutes after the start of the incident. The relaxation of cohort FIFO discipline produced a slightly more gradual convergence to the flows of 900 vph.

6.3.4 Effects on the intersection network

All queues in the intersection network were caused by oversaturated movements. Streams were constrained by the discharge capacity of the approach, which was distributed to the streams in proportion to the partial demand. The local supply never constrained the partial flows, even during the incident scenario. So, the queue which formed on any approach consisted of all movements in proportions equal to the demand. Spillover was not present, so the permission of overtaking among cohorts had no impact on the simulation results. Only a small discrepancy was observed due to the instability found during cohort FIFO discipline, which was not present under relaxed FIFO discipline.

6.3 Explicit lanes

The assumption concerning the dynamic distribution of the discharge capacity among competing streams is optimistic in the models presented thus far. In reality, lane striping in conjunction with partial demands determines the distribution of the discharge capacity. The absence of lanes in the cell-transmission mode introduces significant errors in the intersection model of CellNetLoad. The accuracy in modelling spillback, spillover, and FIFO discipline is also limited by the absence of explicit lanes. In reality, spillback should occur only on lanes with permission to send flow to the arc in the forward star with no remaining local supply. Lanes with permission to other arcs only should remain unblocked until the queue from an adjacent lane spills over.

This situation is remedied in CellNetLoad with the introduction of the explicit representation of lanes in the last cell of each arc. This feature enables the specification of reduced lane capacities, instead of reduced arc capacity. Lane permissions are respected, so individual lanes may be blocked by spillback or oversaturation without immediate spillover across all lanes.

6.3.1 Implementation

An option is available in CellNetLoad to model explicit lanes in CellNetLoad by including the file name of the lane striping file as the twentieth file name in the master file. The format of the lane striping file differs from that of INTEGRATION. All arcs must be included on a separate line of the text file. The first integer is the arc identification number, which is followed by a bitstring for each lane, starting with the median lane. The bitstring contains one character for each arc in the forward star of the node at the downstream end of the arc of this line in the text file. The order of the bits is the ascending order of the arc identification numbers of the arcs in the forward star. A bit value of 1 implies that the corresponding lane has permission to send flow to the corresponding arc in the forward star. A bit value of 0 denies permission for the corresponding stream to use the corresponding lane. The values of the initial flow capacity factors for each lane may be appended to each line, starting with the median lane.

Two additional classes were necessary for the implementation of this enhancement. The LastCell class has a list in which to store LaneCell objects, the first new class. Both the LastCell class and the LaneCell class are subclasses of the Cell class. The LaneCell object contains the lane permissions and an array of cohort lists, one list for each arc in the forward star. The cohort list temporarily stores the packets ready to be discharged from that lane to the first cell of the arc in the forward star.

The Node class also has two subclasses. All variables are stored in the Node class, sufficiently generalised to implement all intersection algorithms. The NodeNoLane subclass contains the procedures which implement the intersection algorithms discussed previously. The NodeLane subclass contains the procedures which implement the intersection algorithms with explicit lanes.

Two options are available for modelling the queue spillover across lanes during congested conditions. The choice is specified with a flag on the command line. One option is the assumption that the queue will remain in the lane(s) with discharge permission, and will never spill over to other lanes. The algorithm for this option is shown in Figure 6.12. A packet may be discharged from an unblocked lane no matter how far back it may be in the cohort queue of the last cell, which is similar to the algorithm without explicit lanes in Figure 6.7.

Figure 6.12 Cell FIFO algorithm for the determination of intersection flow
– overtaking permitted among packets in the same cell
– explicit lanes with lane striping permission on each approach

```
Step 0. Initialisation
        iteration $m = 0$
        vehicle movement $y_{ij\ell}^{m} = 0 \;\forall i \in A^{-}, \ell \in L_{i}, j \in A^{+}$
        local demand remainder $\overline{D}_{i}^{m} = D_{i} \;\forall i \in A^{-}$
        discharge capacity remainder $\overline{Q}_{i\ell}^{m} = Q_{i\ell} \;\forall i \in A^{-}, \ell \in L_{i}$
        local supply remainder $\overline{S}_{j}^{m} = S_{j} \;\forall j \in A^{+}$
        lane availability $\varphi_{i\ell}^{m} = 1 \;\forall i \in A^{-}, \ell \in L_{i}$
        current cohort $\overline{\tau}_{i} = \min_{\tau : n_{i\ell\tau}^{d} > 0}(\tau) \;\forall i \in A^{-}$
```

packet groups to arcs in forward star $n^{m}_{ij\tau_i} = \sum_{d} n^{d}_{ij\tau_i} \quad \forall i \in A^{-}, j \in A^{+}$

Step 1. Calculate partial demand

Step 1a. Determine the feasible partial demand

$$n^{m}_{ij} = \begin{cases} n^{m}_{ij\tau_i} & \text{if } \overline{S}^{m}_{j} > 0 \text{ and } \sum\limits_{\ell \in L_i} \varphi^{m}_{i\ell}\rho_{ij\ell} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^{-}, \; j \in A^{+}$$

$$\forall i \in A^{-}: \qquad \text{if } \sum_{j} n^{m}_{ij} > \overline{D}^{m}_{i} \text{ then } n^{m}_{ij} \leftarrow n^{m}_{ij} / \overline{D}^{m}_{i} \quad \forall j \in A^{+}$$

If $\exists i \in A^{-} : \sum_{j} n^{m}_{ij} = 0$ then continue to step 1b. Otherwise, skip to step 2.

Step 1b. Update the current cohort

$$\forall i \in A^{-}: \qquad \text{if } \sum_{j} n^{m}_{ij} = 0 \text{ and } \overline{\tau}_i + \Delta t \le \max_{\tau : n_{ij\tau} > 0}(\tau) \text{ then}$$

$$\varphi^{m}_{i\ell} \leftarrow \begin{cases} 0 & \text{if } \exists j : \overline{S}^{m}_{j} = 0, n^{m}_{ij\tau_i}\rho_{ij\ell} > 0 \\ \varphi^{m}_{i\ell} & \text{otherwise} \end{cases} \forall \ell \in L_i \, .$$

$$\overline{\tau}_i \leftarrow \overline{\tau}_i + \Delta t \, . \; n^{m}_{ij\tau_i} = \sum_{d} n^{d}_{ij\tau_i} \quad \forall j \in A^{+}$$

Return to step 1a.

Step 2. Calculate the number of vehicles $\Delta y^{m}_{ij\ell}$ to depart lane ℓ of cell i for cell j

$$\Delta y^{m}_{ij} = \min\left(n^{m}_{ij} \Big/ \sum_{\ell \in L_i} \varphi^{m}_{i\ell}\rho_{ij\ell} \, , \alpha^{m}_{ij\ell}\overline{Q}^{m}_{i\ell} \, , \beta^{m}_{ij\ell}\overline{S}^{m}_{j} \right) \quad \forall i \in A^{-}, j \in A^{+}$$

$$\text{where} \quad \alpha^{m}_{ij\ell} = \begin{cases} \varphi^{m}_{i\ell} \Big/ \sum\limits_{j : n^{m}_{ij} > 0} \rho_{ij\ell} & \text{if } \sum\limits_{j : n^{m}_{ij} > 0} \rho_{ij\ell} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^{-}, \ell \in L_i, \; j \in A^{+}$$

$$\beta^{m}_{ij\ell} = \begin{cases} \alpha^{m}_{ij\ell}\overline{Q}^{m}_{i\ell} \Big/ \sum\limits_{i}\sum\limits_{\ell \in L_i} \alpha^{m}_{ij\ell}\overline{Q}^{m}_{i\ell} & \text{if } \alpha^{m}_{ij\ell} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^{-}, \ell \in L_i, \; j \in A^{+}$$

Step 3. Stopping criterion

If $\exists i \in A^{-}, \ell \in L_i, j \in A^{+} : \Delta y^{m}_{ij\ell} > 0$, then continue to Step 4. Otherwise, stop.

Step 4. Update

$$y^{m+1}_{ij\ell} = y^{m}_{ij\ell} + \Delta y^{m}_{ij\ell} \quad \forall i \in A^{-}, \ell \in L_i, j \in A^{+}$$

$$n^{d}_{ij\tau_i} \leftarrow n^{d}_{ij\tau_i} \sum_{\ell \in L_i} \Delta y^{m}_{ij\ell} \Big/ \sum_{d} n^{d}_{ij\tau_i} \quad \forall i \in A^{-}, j \in A^{+}$$

$$n^{m+1}_{ij\tau_i} = n^{m}_{ij\tau_i} - \Delta y^{m}_{ij\ell} \quad \forall i \in A^{-}, j \in A^{+}$$

$$\overline{D}_i^{m+1} = D_i - \sum_{\ell \in L_i} \sum_j y_{ij\ell}^{m+1} \quad \forall i \in A^-$$

$$\overline{Q}_{i\ell}^{m+1} = Q_{i\ell} - \sum_j y_{ij\ell}^{m+1} \quad \forall i \in A^-, \ell \in L_i$$

$$\overline{S}_j^{m+1} = S_j - \sum_i \sum_{\ell \in L_i} y_{ij\ell}^{m+1} \quad \forall j \in A^+$$

$$\varphi_{i\ell}^{m+1} = \left\{ \begin{array}{ll} 0 & \text{if } \exists j : \overline{S}_j^{m+1} = 0, n_{ij\tau_i}^{m+1} \rho_{ij\ell} > 0 \\ 0 & \text{if } \exists j : \overline{Q}_{i\ell}^{m+1} = 0 \\ \varphi_{i\ell}^m & \text{otherwise} \end{array} \right\} \forall i \in A^-, \ell \in L_i$$

$$m \leftarrow m + 1$$

Return to Step 1.

The intersection model was revamped, though care was taken to assure that the data structures remained compatible with the algorithms described earlier. Some variables were modified, new variables were added, and the algorithms were altered.

Several variables were generalised to account for each lane. The lane discharge capacity $Q_{i\ell} = q_{\max,i} \Delta t$ is stored for use in calculations. The distribution coefficient $\alpha_{ij\ell}^m$ allocates a portion of the remaining discharge capacity of lane ℓ in cell i to the stream headed to cell j during iteration m of the algorithm. The distribution coefficient $\beta_{ij\ell}^m$ allocates a portion of the remaining local supply of cell j to lane ℓ of cell i during iteration m of the algorithm. Vehicle movement $y_{ij\ell}^m$ is the total number of vehicles which are prepared to exit lane ℓ of cell i and enter cell j at the m^{th} iteration of the algorithm, whereas $\Delta y_{ij\ell}^m$ represents the quantity which is added to $y_{ij\ell}^m$ during the m^{th} iteration of the algorithm.

Several new variables were added to account for lane striping and blockages. Lane permission

$$\rho_{ij\ell} = \begin{cases} 1 \text{ if lane } \ell \text{ of cell i has permission to send flow to cell j} \\ 0 \text{ otherwise} \end{cases} \quad \forall i \in A^-, \ell \in L_i, \; j \in A^+$$

is read from the lane striping input file and remains constant throughout the simulation. Lane availability

$$\varphi^{mt}_{ij\ell} = \begin{cases} 0 \text{ if lane } \ell \text{ of cell i is blocked} \\ 1 \text{ otherwise} \end{cases} \quad \forall i \in A^-, \ell \in L_i, \; j \in A^+$$

is reset to 1 at the beginning of each time step, then updated during the algorithm to detect spillback or oversaturation.

The cohorts are processed sequentially so that lane blockages may be detected. The cohort, which is currently being used to determine the partial demand, has arc entrance time $\bar{\tau}_i$. The partial demands are determined in step 1 of the algorithm in Figure 6.12. In step 1a only packets with permission on an unblocked lane to the downstream arc with positive remaining local supply are considered for the partial demand. If no packets in the current cohort satisfy these criteria, then the next cohort in the cohort queue is considered. Prior to incrementing the current cohort, each lane is checked for blockage due to spillback in step 1b.

The lane flows are determined in step 2. The first constraint is the partial demand, which is evenly distributed on the unblocked lanes with the appropriate permission. The second constraint evenly distributes the remaining lane discharge capacity among the competing streams with permission and positive partial demand. All streams sharing an unblocked lane are assigned equal portions of the remaining lane discharge capacity regardless of the partial demands of the streams. Any residual lane discharge capacity unused by one stream is reassigned to other streams in subsequent iterations. The third constraint distributes the local supply of the first cell of each arc in the forward star to each lane of each arc in the backward star according to the fraction of the discharge capacity assigned to that stream in that lane. This distribution strategy is weighted by the stream capacity and not by partial demand. Though if the partial demand for a particular stream is zero, then that stream is not allocated a portion of the lane discharge capacity or a portion of the local supply. This greatly improves the efficiency of the algorithm by reducing the unused residual lane discharge capacity and unused residual local supply.

The algorithm is stopped in step 3 when all stream discharges are zero because the partial demand, lane discharge capacity, and/or local supply is exhausted for each stream, or because all lanes with permission for a given stream are blocked by spillback.

The variables are updated in step 4. The partial flow and the occupancy of the current cohort are updated to reflect the movement of packets from the current cohort in the cohort queue to the list of packets ready to depart the cell. Each packet is split evenly among all the lanes with permission to send that packet downstream. The remaining local demand, the remaining lane discharge capacity, and the remaining local supply are updated. Each lane is checked for blockage due to spillback or due to the exhaustion of the lane discharge capacity.

Another algorithm was developed for the alternative model for spillover. In this model it is assumed that those cohorts at the front of the cohort queue which represent the local demand will spill over across all lanes if they are unable to discharge during this time step. This spillover prevents the discharge of any packets in the cohort queue which are farther back than the cohorts which represent the local demand. This is similar to the algorithm without explicit lanes in Figure 6.8.

This alternative algorithm is the same as Figure 6.12, except for step 1 and step 4. The modifications are included in Figure 6.13. The current cohort is split in step 1a if it exceeds the remainder of the local demand. The same restrictions are applied to the feasible partial demand in step 1b. If there is no feasible partial demand in the current cohort, then the current cohort is incremented only if some local demand remains. The only change in step 4 is made to the local demand remainder. The occupancy of all cohorts ahead of the current cohort in the cohort queue is subtracted from the local demand. This prevents the algorithm from considering more cohorts in the cohort queue than those at the front which represent the local demand.

The proof of convergence of for the algorithms with explicit lanes relies on the fact that the partial demands, the remaining discharge capacity, and the remaining local supply are non-increasing. During each iteration, flow is added to at least one stream. Therefore, the partial demand, the remaining discharge capacity, and the remaining local supply of that stream must decrease. The number of iterations is also finite, since the stopping criterion in step 3 is implemented with some small positive value—the default is 0.001 vehicles. The proof of the finiteness of the number of iterations does not apply to the algorithms with explicit lanes because the distribution factors are not weighted by the partial demands.

Figure 6.13 Relaxed cohort FIFO algorithm for the determination of intersection flow
– overtaking permitted among the packets at the front of the cohort queue which represent the local demand
– explicit lanes with lane striping permission on each approach

```
Step 0. Initialisation
        iteration $m = 0$
        vehicle movement $y_{ij\ell}^{m} = 0 \ \forall i \in A^{-}, \ell \in L_i, j \in A^{+}$
        local demand remainder $\overline{D}_i^{m} = D_i \ \forall i \in A^{-}$
        discharge capacity remainder $\overline{Q}_{i\ell}^{m} = Q_{i\ell} \ \forall i \in A^{-}, \ell \in L_i$
        local supply remainder $\overline{S}_j^{m} = S_j \ \forall j \in A^{+}$
        lane availability $\varphi_{i\ell}^{m} = 1 \ \forall i \in A^{-}, \ell \in L_i$
        current cohort $\overline{\tau}_i = \min_{\tau : n_{ij\tau}^{d} > 0}(\tau) \ \forall i \in A^{-}$
        packet groups to arcs in forward star $n_{ij\overline{\tau}_i}^{m} = \sum_d n_{ij\overline{\tau}_i}^{d} \ \forall i \in A^{-}, j \in A^{+}$
Step 1. Calculate partial demand
```

Step 1a. Check that the local demand is separate from the rest of the cohort queue

$\forall i \in A^-$: if $\sum_j n_{ij\bar{\tau}}^m > \bar{D}_i^m$ then split the cohort with time stamp $\bar{\tau}_i$

into two cohorts with time stamps $\bar{\tau}_{i1} = \bar{\tau}_i$ and $\bar{\tau}_{i2} = \bar{\tau}_i + \varepsilon$

such that $\sum_j n_{ij\bar{\tau}_{i1}}^m = \bar{D}_i^m$

Step 1b. Determine the feasible partial demand

$$n_{ij}^m = \begin{cases} n_{ij\bar{\tau}}^m & \text{if } \bar{S}_j^m > 0 \text{ and } \sum_{l \in L_i} \varphi_{il}^m \rho_{ijl} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^-, \; j \in A^+$$

$\forall i \in A^-$: if $\sum_j n_{ij}^m > \bar{D}_i^m$ then $n_{ij}^m \leftarrow n_{ij}^m / \bar{D}_i^m$ $\forall j \in A^+$

If $\exists i \in A^-$: $\sum_j n_{ij}^m = 0$ then continue to step 1c. Otherwise, skip to step 2.

Step 1c. Update the current cohort

$\forall i \in A^-$: if $\sum_j n_{ij}^m = 0$ and $\bar{\tau}_i + \Delta t \le \max_{\tau: n_{ij\tau}^m > 0}(\tau)$ and $\sum_j n_{ij\bar{\tau}}^m < \bar{D}_i^m$ then

$$\varphi_{il}^m \leftarrow \begin{cases} 0 & \text{if } \exists j : \bar{S}_j^m = 0, n_{ij\bar{\tau}}^m \rho_{ijl} > 0 \\ \varphi_{il}^m & \text{otherwise} \end{cases} \quad \forall l \in L_i,$$

$\bar{\tau}_i \leftarrow \bar{\tau}_i + \Delta t$, $n_{ij\bar{\tau}}^m = \sum_d n_{ij\bar{\tau}}^d$ $\forall j \in A^+$

Return to step 1a

Step 2. Calculate the number of vehicles Δy_{ijl}^m to depart lane l of cell i for cell j

$$\Delta y_{ijl}^m = \min\left(n_{ij}^m \Big/ \sum_{l \in L_i} \varphi_{il}^m \rho_{ijl}, \alpha_{ijl}^m \bar{Q}_{il}^m, \beta_{ijl}^m \bar{S}_j^m \right) \quad \forall i \in A^-, j \in A^+$$

$$\text{where} \quad \alpha_{ijl}^m = \begin{cases} \varphi_{il}^m \Big/ \sum_{j: n_{ij}^m > 0} \rho_{ijl} & \text{if } \sum_{j: n_{ij}^m > 0} \rho_{ijl} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^-, l \in L_i, \; j \in A^+$$

$$\beta_{ijl}^m = \begin{cases} \alpha_{ijl}^m \bar{Q}_{il}^m \Big/ \sum_i \sum_{l \in L_i} \alpha_{ijl}^m \bar{Q}_{il}^m & \text{if } \alpha_{ijl}^m > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in A^-, l \in L_i, \; j \in A^+$$

Step 3. Stopping criterion

If $\exists i \in A^-, l \in L_i, j \in A^+ : \Delta y_{ijl}^m > 0$, then continue to Step 4. Otherwise, stop.

Step 4. Update

$y_{ijl}^{m+1} = y_{ijl}^m + \Delta y_{ijl}^m$ $\forall i \in A^-, l \in L_i, j \in A^+$

$$n^{d}_{ij\tau_i} \leftarrow n^{d}_{ij\tau_i} \sum_{\ell \in L_i} \Delta y^{m}_{ij\ell} \Big/ \sum_{d} n^{d}_{ij\tau_i} \quad \forall i \in A^{-}, j \in A^{+}$$

$$n^{m+1}_{ij\tau_i} = n^{m}_{ij\tau_i} - \Delta y^{m}_{ij\ell} \quad \forall i \in A^{-}, j \in A^{+}$$

$$\overline{D}^{m+1}_{i} = D_i - \sum_{\ell \in L_i} \sum_{j} y^{m+1}_{ij\ell} - \sum_{\tau < \tau_i} \sum_{j} \sum_{d} n^{d}_{ij\tau} \quad \forall i \in A^{-}$$

$$\overline{Q}^{m+1}_{i\ell} = Q_{i\ell} - \sum_{j} y^{m+1}_{ij\ell} \quad \forall i \in A^{-}, \ell \in L_i$$

$$\overline{S}^{m+1}_{j} = S_j - \sum_{i} \sum_{\ell \in L_i} y^{m+1}_{ij\ell} \quad \forall j \in A^{+}$$

$$\varphi^{m+1}_{i\ell} = \left\{ \begin{matrix} 0 & \text{if } \exists j : \overline{S}^{m+1}_{j} = 0, n^{m+1}_{ij\tau_i} \rho_{ij\ell} > 0 \\ 0 & \text{if } \exists j : \overline{Q}^{m+1}_{i\ell} = 0 \\ \varphi^{m}_{i\ell} & \text{otherwise} \end{matrix} \right\} \forall i \in A^{-}, \ell \in L_i$$

$$m \leftarrow m + 1$$

Return to Step 1.

The third option which was available for CellNetLoad without explicit lanes assumes that spillover occurs instantaneously across all lanes of the arc. For obvious reasons, this option is not available for CellNetLoad with explicit lanes. In all algorithms, FIFO discipline is respected within a lane and within a stream.

6.4.2 Effects on the diverge network

The results of the CellNetLoad simulation with explicit lanes and overtaking permitted within a cell are shown in Figure 6.14. The saturation flow arrived at the diverge at 940 s. The through vehicles used both lanes, which limited the availability of the right lane for the right turners and causing a small queue of right turners. As the queue of right turners grew, they took priority on the right lane according to FIFO discipline in the lane discharge, and forced the through vehicles to use the left lane only. This sequence of events occurs earlier in the simulation with the 5-s time step. As the density in the last cell grew past 40 veh/km/lane, the local supply of the last cell decreased from 1800 vphpl, causing a queue to build upstream of the last cell, and preventing some of the left turners and through vehicles from reaching the last cell. This also happened earlier in the simulation with the 5-s time step.

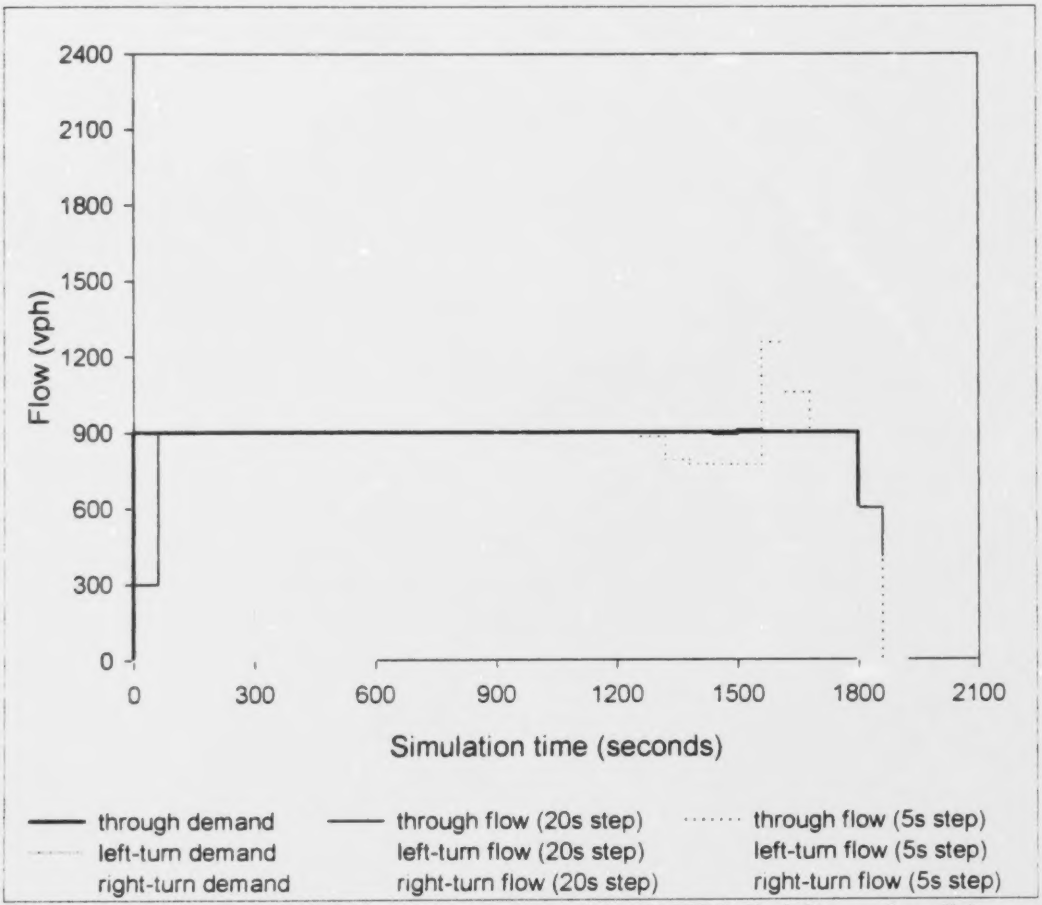


Figure 6.14 Flow from CellNetLoad at a diverge with cell FIFO discipline and explicit lanes

The results of the CellNetLoad simulation with explicit lanes and overtaking permitted at the front of the cohort queue, which represents the local demand, are shown in Figure 6.15. A portion of the right turners in the local demand formed a queue at 960 s. These right turners always passed through the diverge during the next time step, along with some new local demand of through vehicles, left turners, and right turners in the ratio $\frac{900}{3000}:\frac{300}{3000}:\frac{1800}{3000}$. So, the through vehicles and left turners never reached flows of 900 vph and 300 vph, respectively. The through vehicles were always distributed on both lanes, so the right turners never reached 1800 vph in the right lane. At 1240 s, the demand ratios changed to $\frac{900}{3300}:\frac{300}{3300}:\frac{2100}{3300}$, which is reflected in the observed flows. The instabilities in the observed flows were a result of the varying size of the remaining right-turners in the local demand during each time step.

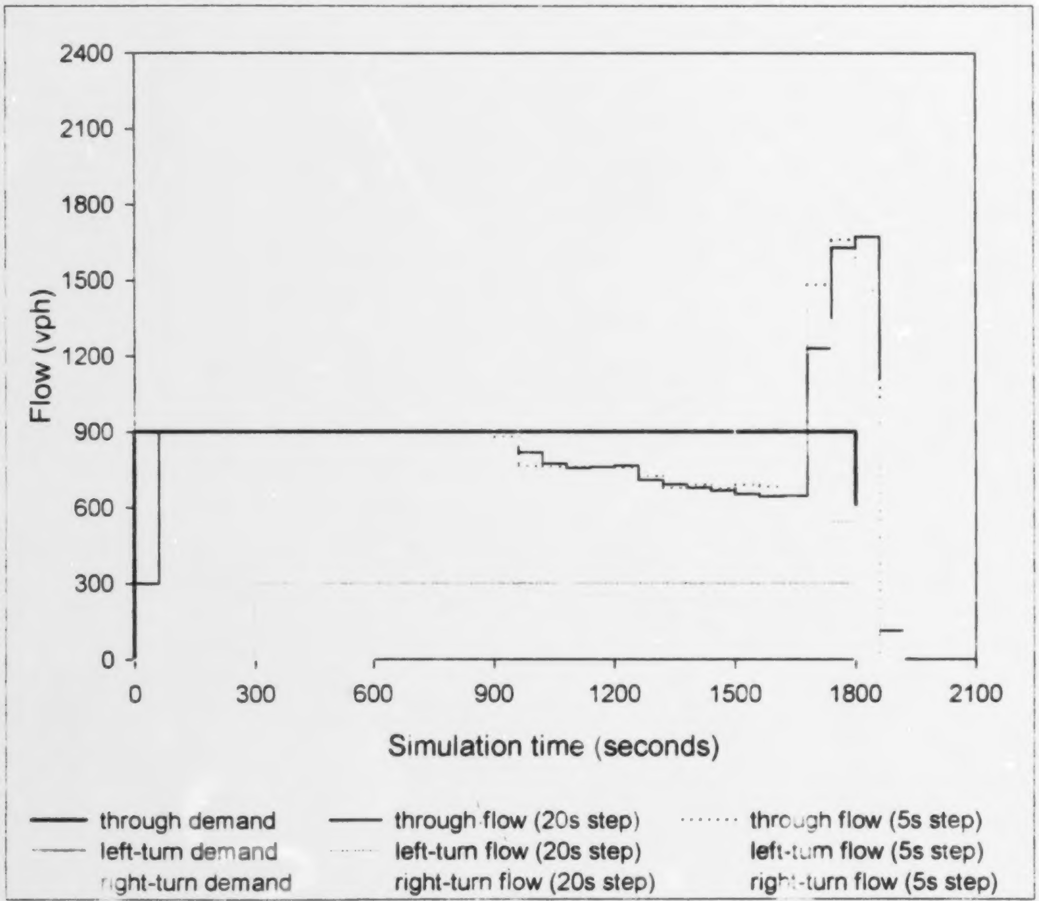


Figure 6.15 Flow from CellNetLoad at a diverge with relaxed cohort FIFO discipline and explicit lanes

Several desirable conclusions can be drawn from the simulation results with explicit lanes. Firstly, the observed flows seem less dependent on the time step. Secondly, the observed flows match the benchmark results shown in Figure 4.17 relatively well.

6.4.3 Effects on the merge network

The arcs in the backward star of the merge node each carry a single stream on a single lane. So, spillover is not a factor. Because all intersection models respect FIFO discipline on individual lanes and for each stream, the relaxation of cohort FIFO discipline has no effect. In fact, the only impact of the intersection algorithms with explicit lanes on this merge network is the different distribution of the local supply during congested conditions. The local supply is distributed evenly to each lane of the arcs in the backward star, so each stream has an equal share of the local supply, instead of weighting the distribution by the partial demands.

The resulting flows into the merge are shown in Figure 6.16. During the first incident from 300 to 600 s, each stream is allocated 900 vph of the local supply, forcing the mainline flow to decrease from 1200 vph. There is some delay in the decrease for the simulation with a 5-s time step because the congestion from the incident must spill back three cells before affecting the local supply of the merge cell. The effect of the second incident from 1200 to 1500 s is opposite to the effect of the first due to the switching of the demands and the symmetry with which lane drops are modelled in CellNetLoad.

6.4.4 Effects on the intersection network

CellNetLoad can accommodate capacity reductions at the exit of specific lanes when explicit lanes are modelled. The lane capacity reductions, which were derived from the INTEGRATION simulation results, were included in the lane striping input file as the initial flow capacity factors. The flow capacity factors are shown in Figure 6.17. Greater capacity reductions were applied to the lanes carrying the left turn movement to account for gap acceptance behaviour. The capacity of the eastbound exclusive through lane is reduced more than the shared through/right-turn lane because of occasional spillover from the left turn pocket.

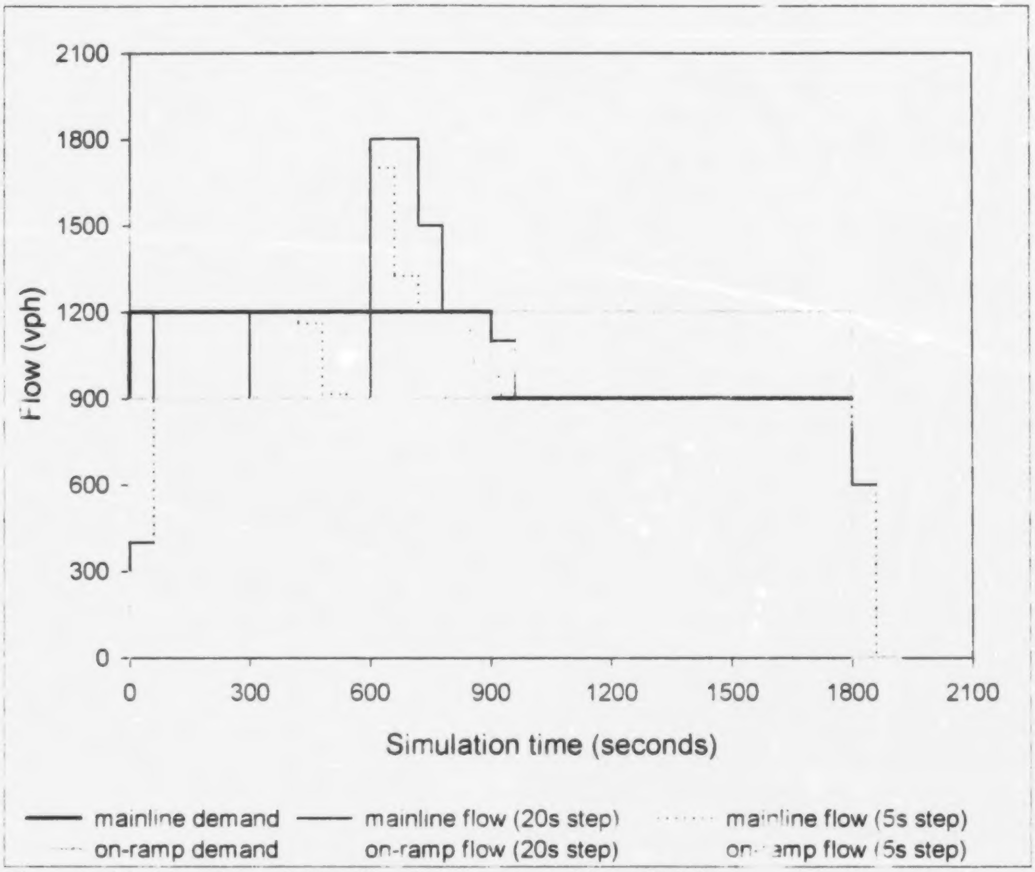


Figure 6.16 Flow from CellNetLoad at a merge with explicit lanes

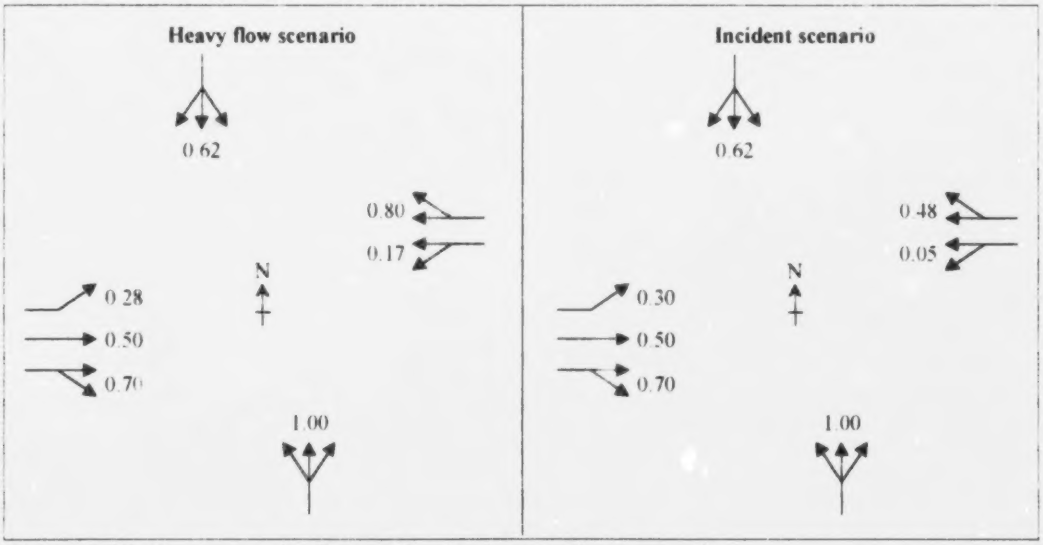
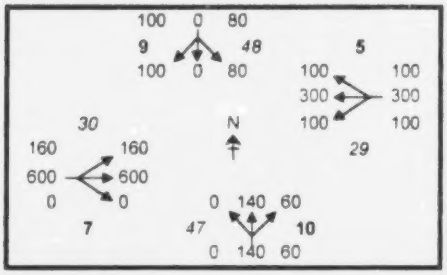

Heavy flow scenario
0.62
0.80
0.17
N
0.28
0.50
0.70
1.00
Incident scenario
0.62
0.48
0.05
N
0.30
0.50
0.70
1.00

Figure 6.17 Lane capacity factors for the intersection network

The CellNetLoad simulation results with cell FIFO discipline are shown in Figure 6.18. The light flow scenario is identical to the simulation results previously shown. In the heavy flow scenario, the oversaturation of the southbound approach, the eastbound left turners, and the westbound left-turners was expected. In the incident scenario, the oversaturation was greater than expected for the right-turn and through movements on the westbound approach due to extensive queuing of the left-turn movement which spilled back to the upstream cell. This spillback mimics the spillover, but is dependent on the cell length, which is undesirable.

Light Flow

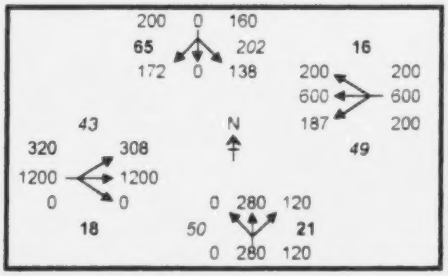


100 0 80
9
48
5
100 0 80
100 100
300 300
100 100
30
N
29
160 160
600 600
0 0
7
47
0 140 60
10
0 140 60

approach	demand (vph) left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph) left	through	right	density (veh/km/ln)	travel time (seconds)
E bound	160	600	0	**6.5**	*30*
W bound	100	300	100	**5.1**	*29*
S bound	80	0	100	**9.0**	*48*
N bound	0	140	60	**9.7**	*47*

Heavy flow

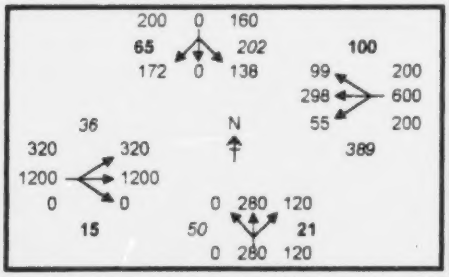


200 0 160
65
202
16
172 0 138
200 200
600 600
187 200
43
N
49
320 308
1200 1200
0 0
18
50
0 280 120
21
0 280 120

approach	demand (vph) left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph) left	through	right	density (veh/km/ln)	travel time (seconds)
E bound	308	1200	0	**17.8**	*43*
W bound	187	600	200	**16.4**	*49*
S bound	138	0	172	**65.0**	*202*
N bound	0	280	120	**20.9**	*50*

Heavy flow with an incident



200 0 160
65
202
100
172 0 138
99 200
298 600
55 200
36
N
389
320 320
1200 1200
0 0
15
50
0 280 120
21
0 280 120

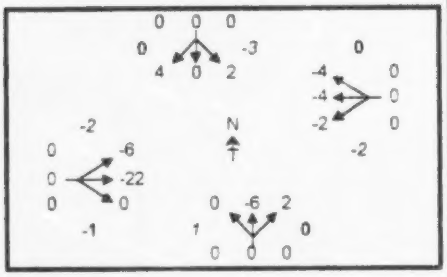
approach	demand (vph) left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph) left	through	right	density (veh/km/ln)	travel time (seconds)
E bound	320	1200	0	**15.2**	*36*
W bound	55	298	99	**99.6**	*389*
S bound	138	0	172	**65.0**	*202*
N bound	0	280	120	**20.9**	*50*

Figure 6.18 Simulation results from CellNetLoad of an intersection with cell FIFO discipline and explicit lanes

The INTEGRATION simulation results were subtracted from the CellNetLoad simulation results for comparison in Figure 6.19. The observed flows match the benchmark closely except for the westbound through movement and right-turn movement during the incident scenario. This was caused by the double penalisation for the spillover of left turners: once in the capacity reduction, and again in the spillback to the upstream cell which caused the right turners and through vehicles to join the queue. The density and travel time are generally underestimated by CellNetLoad compared to the benchmark, even if the outflow matches for all movements on the approach.

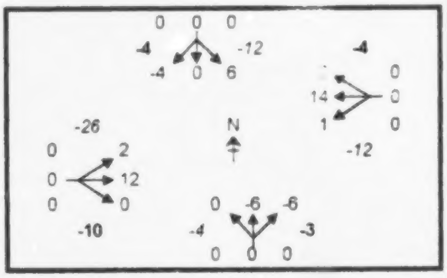
Light Flow



approach	demand (vph)		
	left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	-6	-22	0	-0.7	-2
W bound	-2	-4	-4	-0.3	-2
S bound	2	0	4	-0.4	-3
N bound	0	-6	2	-0.2	1

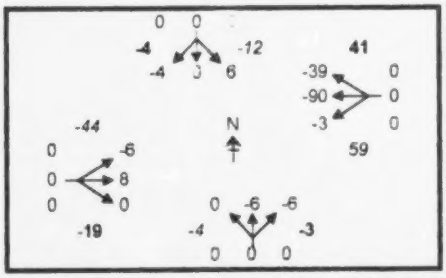
Heavy flow



approach	demand (vph)		
	left	through	
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	2	12	0	-10.4	-26
W bound	1	14	2	-4.3	-12
S bound	6	0	-4	-4.3	-12
N bound	0	-6	-6	-3.0	-4

Heavy flow with an incident



approach	demand (vph)		
	left	through	right
E bound	0	0	0
W bound	0	0	0
S bound	0	0	0
N bound	0	0	0

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	-6	8	0	-19.5	-44
W bound	-3	-90	-39	41.4	59
S bound	6	0	-4	-4.3	-12
N bound	0	-6	-6	-3.0	-4

Figure 6.19 Simulation results from CellNetLoad of an intersection with cell FIFO discipline and explicit lanes compared to INTEGRATION

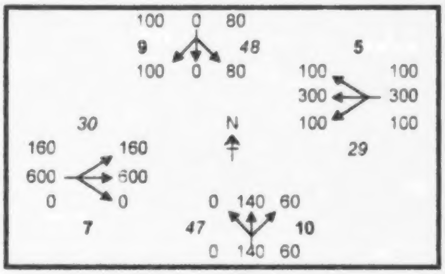
The CellNetLoad simulation results with relaxed cohort FIFO discipline are shown in Figure 6.20. The double penalisation for the spillover is more severe in this model. The southbound approach is identical to the results for cell FIFO discipline because no spillover can occur on a single-lane approach. However, the eastbound through movement shows some evidence of spillover from the left turners in the heavy flow scenario. The same is true for the westbound right-turners and through vehicles on the westbound approach during heavy flow. The double penalisation is more severe on the westbound approach in the incident scenario. The double penalisation creates a greater discrepancy between the CellNetLoad simulation results and the benchmark for the case of relaxed cohort FIFO.

The capacity reductions could account for the double penalisation, but the model which generates the capacity reductions would have to be even more complex. It can be concluded that accurately reproducing the capacity reductions is essential to arriving at accurate predictions of the flow, density and travel time. The cell-transmission model is not well suited to predict the capacity reduction caused by lane-changing and by opposed permissive movements at intersections.

6.5 Effects on the simulation execution time

The simulation time is one component of the total execution time of CellNetLoad. The initialisation, input-processing, and post-processing are excluded from this discussion because they are a reflection of the *implementation* of the cell-transmission model, and not a measure of the performance of the cell-transmission model itself.

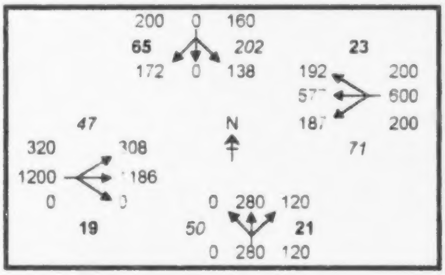
Light Flow


100
0
80
9
48
100
0
80
5
100
100
300
300
100
100
30
N
29
160
160
600
600
0
0
0
140
60
7
47
10
0
140
60

approach	demand (vph)		
	left	through	right
E bound	160	600	0
W bound	100	300	100
S bound	80	0	100
N bound	0	140	60

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	160	600	0	6.5	*30*
W bound	100	300	100	5.1	*29*
S bound	80	0	100	9.0	*48*
N bound	0	140	60	9.7	*47*

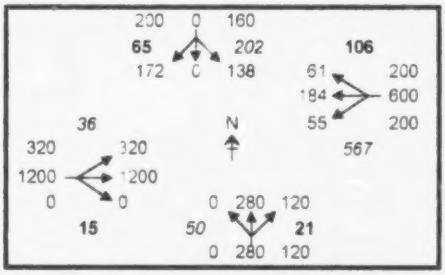
Heavy flow


200
0
160
65
202
23
172
0
138
192
200
577
600
187
200
47
N
71
320
308
1200
1186
0
0
0
280
120
19
50
21
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	308	1186	0	**18.9**	*47*
W bound	187	577	192	**23.2**	*71*
S bound	138	0	172	**65.0**	*202*
N bound	0	280	120	**20.9**	*50*

Heavy flow with an incident


200
0
160
65
202
106
172
0
138
61
200
184
600
55
200
36
N
567
320
320
1200
1200
0
0
0
280
120
15
50
21
0
280
120

approach	demand (vph)		
	left	through	right
E bound	320	1200	0
W bound	200	600	200
S bound	160	0	200
N bound	0	280	120

approach	volume (vph)			density	travel time
	left	through	right	(veh/km/ln)	(seconds)
E bound	320	1200	0	**15.2**	*36*
W bound	55	184	61	**106.3**	*567*
S bound	138	0	172	**65.0**	*202*
N bound	0	280	120	**20.9**	*50*

Figure 6.20 Simulation results from CellNetLoad of an intersection with relaxed cohort FIFO discipline and explicit lanes

It was found in chapter 5 that the simulation execution time is $O\left((\bar{t}/\Delta t)^2\right)$, where the simulation duration $\bar{t}$ is divided into time steps of length Δt. Excluding any effects of the relaxation of cohort FIFO discipline and the time required by the generalised intersection model, the simulation execution time can be estimated in seconds as

$$[6.3] \quad 0.311+\left(7.55\times10^{-5}+5.81\times10^{-5}\|\mathrm{C}\|\right)\left(\bar{t}/\Delta t\right)+2.58\times10^{-4}\left(\|\mathrm{A}\|+\|\mathrm{O}\|\right)\left(1+\bar{t}/\Delta t_s\right)$$

where $\|C\|$, $\|A\|$, and $\|O\|$ are the number of cells, arcs, and origins, respectively; and Δt_s is the output statistics interval. The first term is the overhead. The second term accounts for the clock tick events. The third term accounts for the output statistics events. The coefficients were derived from regression analysis in chapter 5.

6.5.1 Simulation execution time of the diverge network

The predicted simulation execution time from [6.3] for the diverge network is shown as a solid line in Figure 6.21. Five different algorithms were used to simulate the diverge network with three time steps each. The simulation execution times are shown in Figure 6.21. All the algorithms respect FIFO discipline within a traffic stream. FIFO discipline is also respected on each lane, for algorithms which consider lanes explicitly.

The basic algorithm Figure 3.7, called cohort FIFO, permits overtaking of one packet over another if both packets are in the same cohort. Lanes are not modelled explicitly. Spillover occurs across all lanes as soon as spillback is present in a cell. The simulation execution time is labelled with unfilled circles in Figure 6.21. The observed values exceed the predicted values due to the extra time required by the generalised intersection algorithm. The number of iterations of the algorithm depends on the flows. It appears that fewer iterations are required for smaller time steps. This may be explained by the smaller number of vehicles sent during each time step while the vehicle unit, which is used in the stopping criterion, remains constant.

The simulation execution times for the cell FIFO algorithm (Figure 6.7) are shown as unfilled squares in Figure 6.21. Overtaking is permitted within a cell. The implementation of this algorithm requires the traversal of the cohort queue in the last cell of the arcs in the backward star of the node. The number of cohorts in a cell depends on the level of congestion. During uncongested conditions, the cohort queue contains a single cohort. An increase in congestion leads to more cohorts in the cohort queue, and hence an increase in the simulation execution time. From Figure 6.9, it can be seen that the congestion increased for smaller time steps, which explains the increasing gap between the unfilled circles and unfilled squares in Figure 6.21.

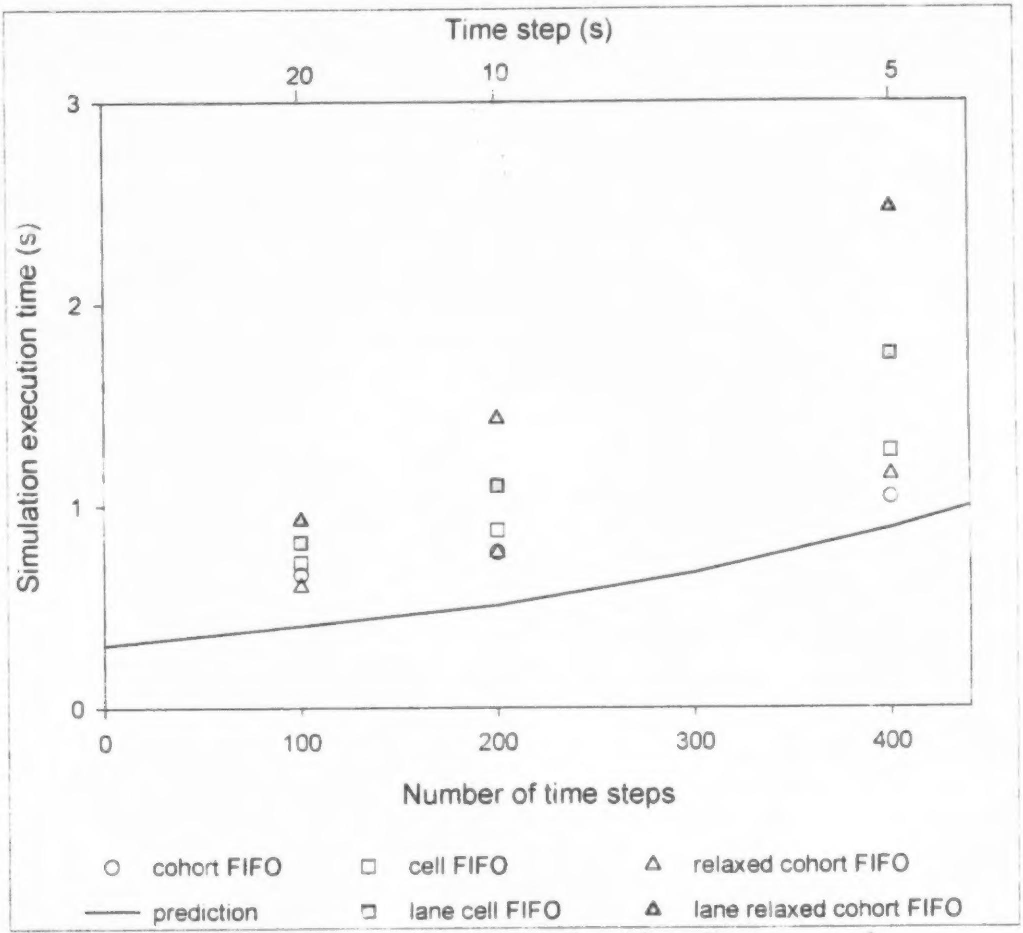


Figure 6.21 Simulation execution time of CellNetLoad for a diverge network

The simulation execution times for the relaxed cohort FIFO algorithm (Figure 6.8) are shown as unfilled triangles in Figure 6.21. Overtaking is permitted among those cohort(s), which represent the local demand, at the front of the cohort queue. The implementation of this algorithm requires the traversal of only the front of the cohort queue in the last cell of the arcs in the backward star of the node. From Figure 6.21, it can be seen that the amount of congestion is almost the same for all time steps. Despite having to traverse multiple cohorts during congested conditions, the simulation execution time is smaller than that of the cohort FIFO algorithm for large time steps. This is explained by a faster convergence, hence fewer iterations, of the intersection algorithm. The convergence is slower for small time steps, since the triangle is above the circle in Figure 6.21 for the simulation with 400 time steps.

Both of the algorithms with explicit lanes had longer execution times than the algorithms without explicit lanes. More calculations are required for each iteration of the algorithms with explicit lanes, and the convergence is probably slower due to the different distribution of the discharge capacity and local supply. Recall that streams on a shared lane are allocated equal portions of the lane discharge capacity, instead of weighting the distribution by the partial demand. As shown in Figure 6.10, the lane relaxed FIFO algorithm predicted more severe congestion for a longer duration than the other algorithms. For this reason, the lane relaxed cohort FIFO algorithm, shown as filled triangles in Figure 6.21, has a longer simulation execution time compared to the lane cell FIFO algorithm, shown as filled squares in Figure 6.21.

The simulation execution times of CellNetLoad from the diverge network indicate that the generalised intersection algorithm has a significant impact on the computational efficiency of the cell-transmission model. Therefore, careful consideration should be given to the trade-off between the realism of the simulation results and the program execution time.

6.5.2 Simulation execution time of the merge network

The predicted simulation execution time from [6.3] for the merge network is shown as a solid line in Figure 6.22. The same five intersection algorithms were used to simulate the merge network with three time steps each. The simulation execution times are shown in Figure 6.21. As explained earlier, the relaxation of cohort FIFO discipline has no effect on the simulation results because FIFO discipline is always enforced for each stream and for each lane.

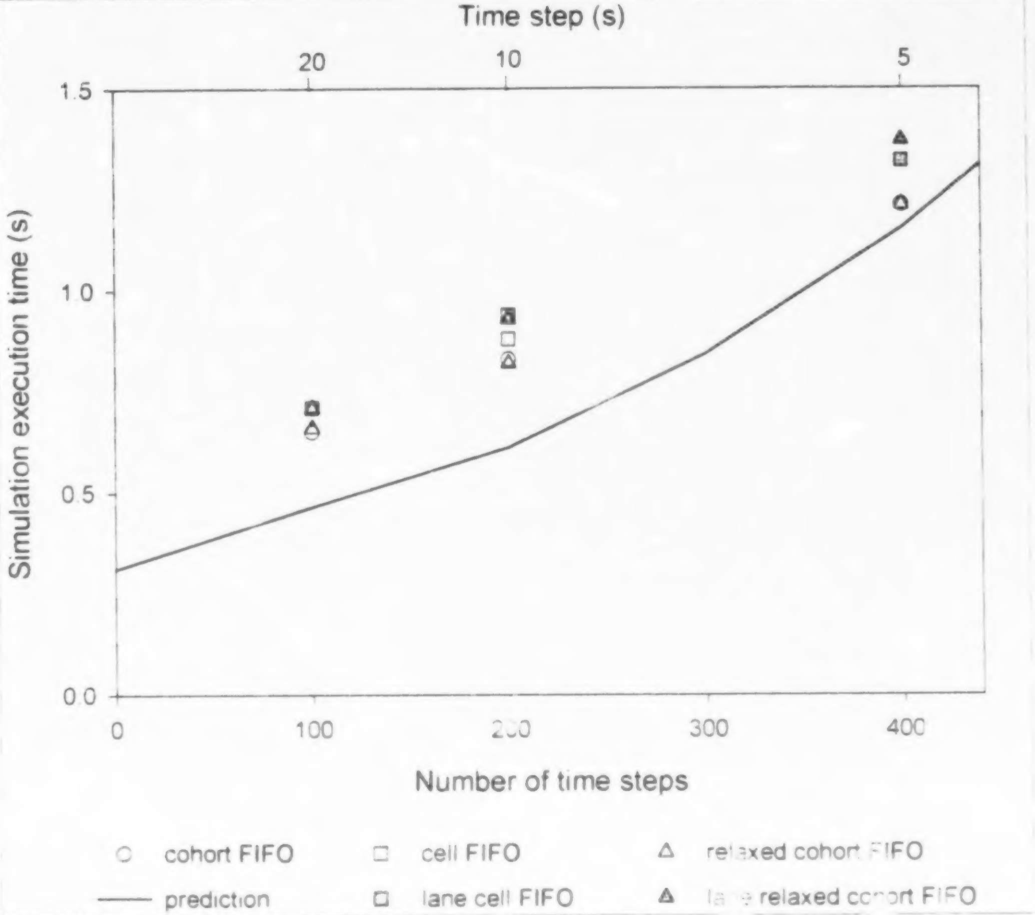


Figure 6.22 Simulation execution time of CellNetLoad for a merge network

The discrepancy between the predicted simulation execution time and that of the cohort FIFO algorithm is attributed to the intersection algorithm. Fewer iterations of the intersection algorithm were required for smaller time steps due to the duration of congestion. The congestion lasted the entire duration of the in300-s incident for the simulations with 20-s and 10-s time steps, which had 100 and 200 time steps per simulation, respectively. The congested was shorter in duration for the simulation with a 5-s time step because the incident did not affect the merge until the congestion spilled back 3 cells to reach the merge cell.

The cell FIFO algorithm, shown with unfilled squares on Figure 6.22, took longer than the cohort FIFO and relaxed cohort FIFO algorithms, shown as unfilled circles and triangles, because it was forced to traverse the entire cohort queue of the last cells of the arcs in the backward star. (The unfilled squares are behind the filled squares for two of the three simulations in Figure 6.22.)

The algorithms with explicit lanes take slightly longer than the other algorithms because they have more overhead than during each iteration. The difference between the algorithm with explicit lanes and those without lanes increases linearly with the number of time steps as expected.

6.5.2 Simulation execution time of the intersection network

The predicted simulation execution time from [6.3] for the intersection network is shown as a solid line in Figure 6.23. The same five intersection algorithms were used to simulate the intersection network with three time steps each for three scenarios. The time steps were 5, 2.5, and 1 s. The first scenario has light flow, the second has heavy flow, and the last has heavy flow and an incident on the eastbound arterial downstream of the intersection. Reduced arc and lane capacities were specified in the input files for the heavy flow scenario and the incident scenario. The predicted simulation execution time is independent of congestion, so the same curve is shown on all figures.

The discrepancy between the predicted and the observed simulation execution times is attributed to the intersection algorithm which is executed once per time step. It was found that the prediction was larger than the observation for the simulation with a 1-s time step. The coefficients for the prediction were derived from regression analysis of the linear network in which all of the cells were last cells. As the time step was decreased, more short arcs were used to represent the linear roadway, so there was only one cell per arc. The intersection network consisted of eight arcs for all of the simulation runs, so there were more regular cells as the time step was decreased. The calculation and movement of flow from a regular cell takes less time than the same operations on a last cell. Thus, the predicted simulation execution time is not reliable for this network. (The merge and diverge networks consisted of last cells only, so the predictions are valid.)

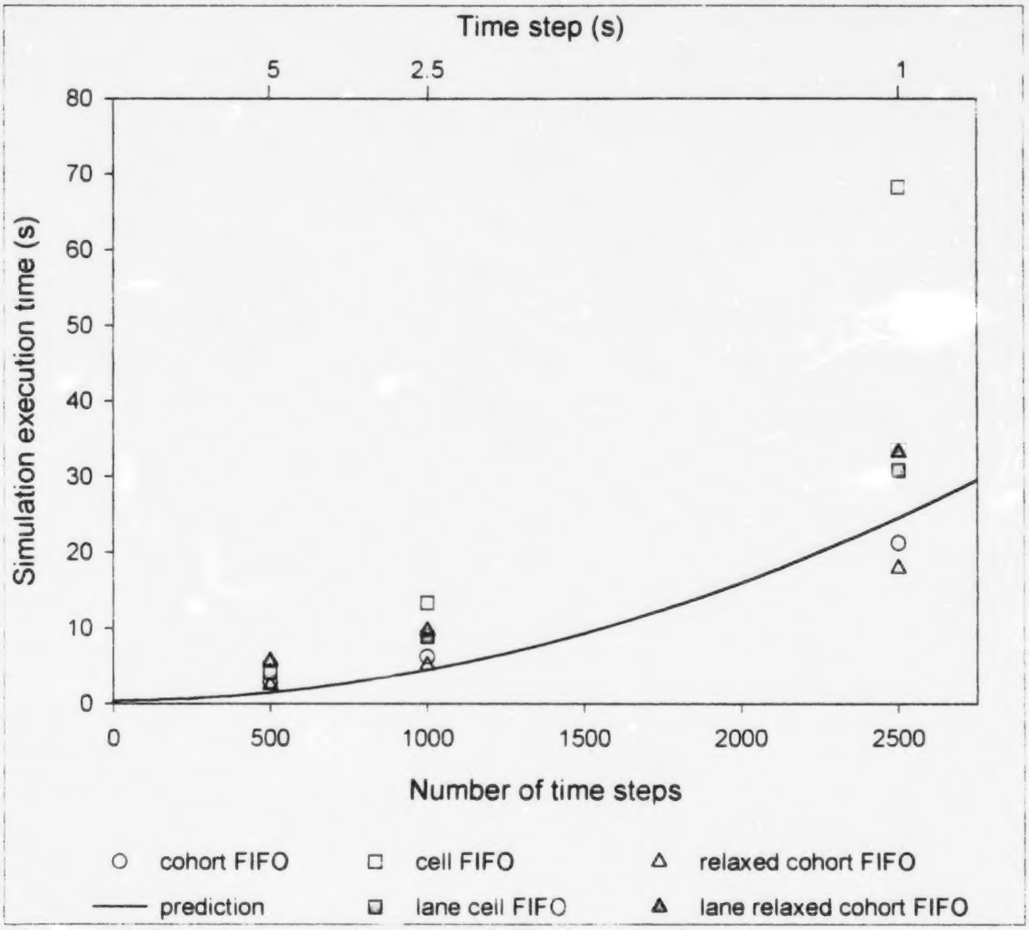


Figure 6.23 Simulation execution time of CellNetLoad for the intersection network with light flow (linear scale)

A log scale was found to be more appropriate for displaying the range of execution times. The simulation execution times for the light flow scenario are shown in Figure 6.24 with a log scale. All congestion during the light flow scenario resulted from the queuing at the signal, since no oversaturation was observed. The relaxed cohort FIFO algorithm converged faster than the cohort FIFO algorithm, while the cell FIFO algorithm converged much more slowly. The algorithms with explicit lanes had nearly the same simulation execution times.

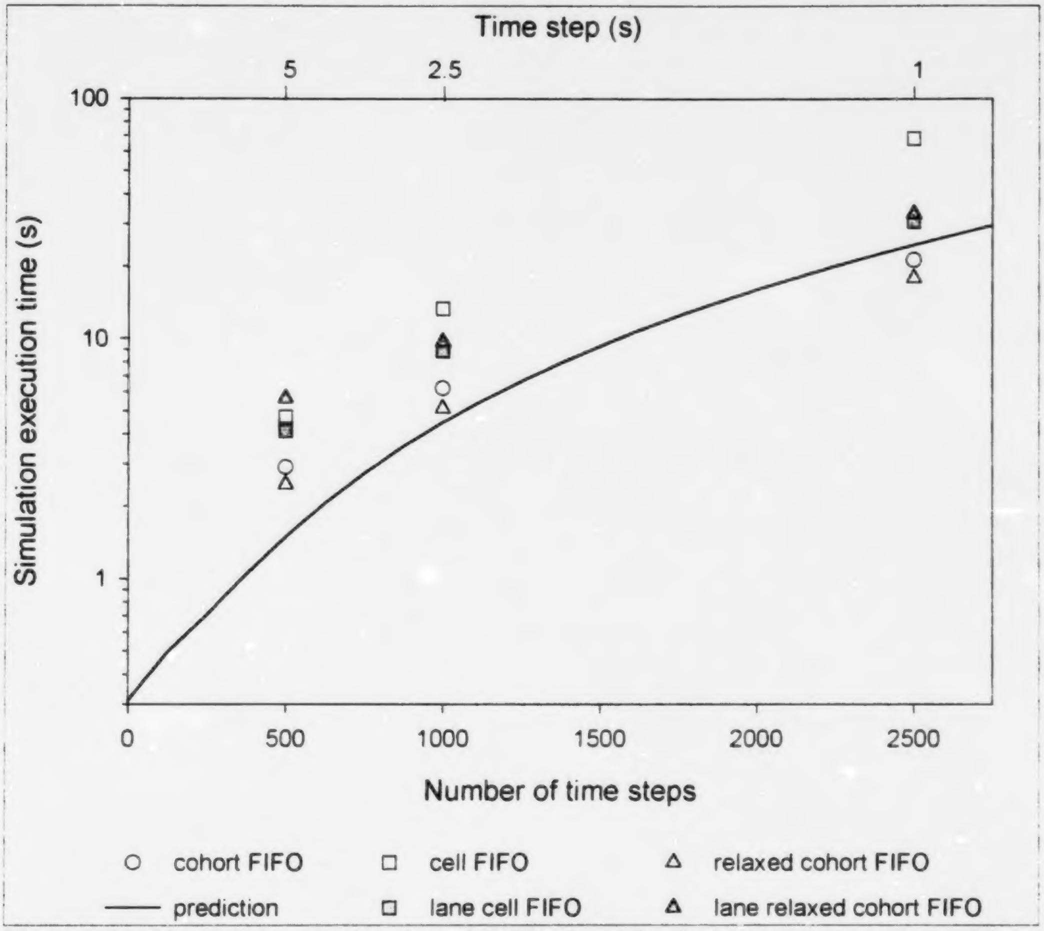


Figure 6.24 Simulation execution time of CellNetLoad for the intersection network with light flow (log scale)

The same data is shown in Figure 6.25 for the heavy flow scenario. Oversaturation caused a great deal of queuing on three of the approaches. The congestion slowed the cell FIFO algorithm and the lane cell FIFO algorithm most. In addition, the algorithms with explicit lanes had slower convergence due to the different distribution of the discharge supply.

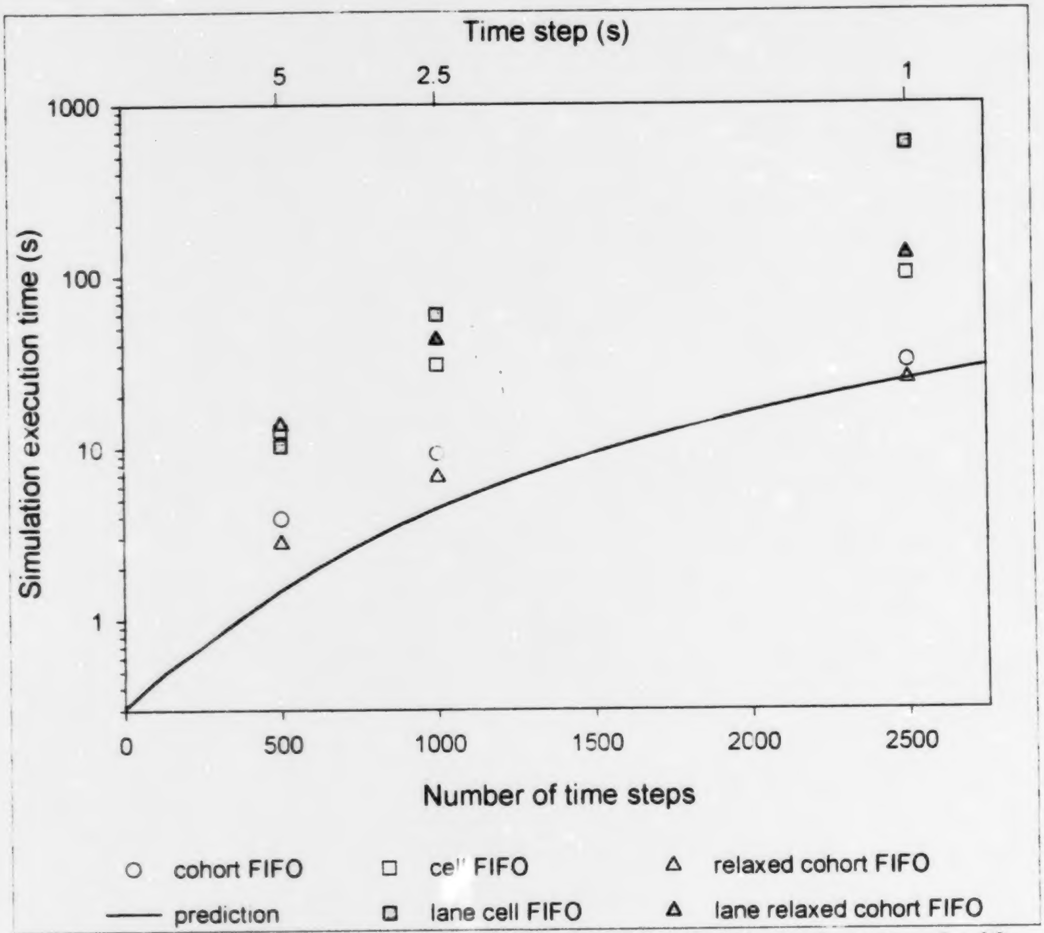


Figure 6.25 Simulation execution time of CellNetLoad for the intersection network with heavy flow

The simulation execution times followed a similar pattern for the heavy flow scenario with the incident (Figure 6.26), but they were larger due to the larger queue on the westbound approach. It is interesting to compare the INTEGRATION simulation duration of 3 minutes for this scenario, compared to the peak of over 80 minutes for CellNetLoad simulation with the lane cell FIFO algorithm and a 1-s time step.

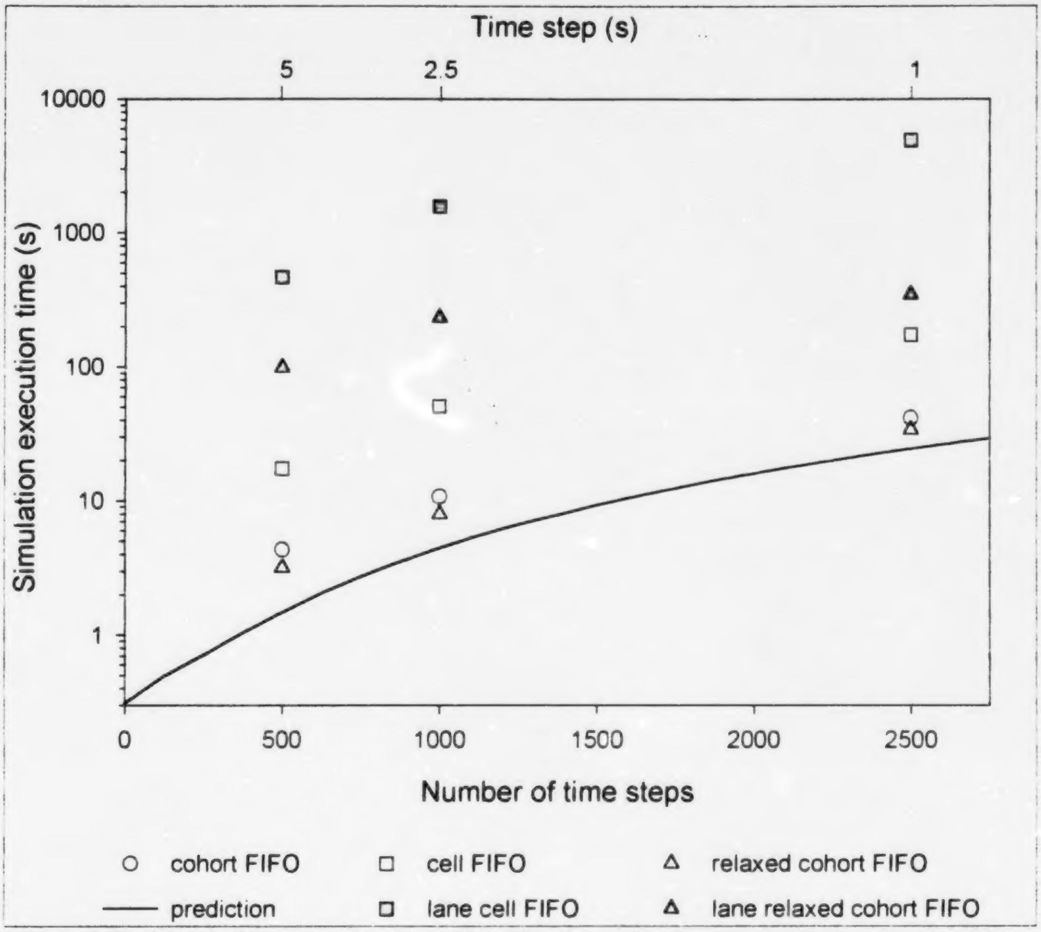


Figure 6.26 Simulation execution time of CellNetLoad for the intersection network with heavy flow and an incident

The intersection network demonstrates the potential for huge simulation execution times of CellNetLoad when trying to test for convergence with the solution of the LWR hydrodynamic model. The simulation results for the intersection network were nearly the same for the 5-s and 1-s time steps.

6.6 Conclusions

The generalised intersection model of CellNetLoad allows realistic representations of intersections, instead of being limited to diverges and merges only. Five algorithms for the intersection model are available in CellNetLoad. The basic algorithm enforces cohort FIFO discipline, which allows overtaking of packets within the same cohort only. As a result, spillover impedes all streams as soon as one stream spills back on the approach. The discharge capacity and local supply are distributed to competing streams in proportion to the partial demand. This algorithm is appropriate for single-lane approaches.

The enhancements to the intersection take two forms. Firstly, the cohort FIFO discipline is relaxed to better represent multi-lane approaches. The relaxed cohort FIFO algorithm permits overtaking among the cohort(s) at the front of the cohort queue which represent the local demand. This delays the effects of spillover until the occupancy of the last cell of the approach exceeds the number of vehicles which may exit at the capacity flow. The cell FIFO algorithm permits overtaking within the last cell of the approach. So, spillover is delayed until the queue spills back to the cell upstream of the last cell on the approach. Neither of these relaxations is appropriate for single-lane approaches. The relaxed cohort FIFO algorithm tends to converge faster than the cohort FIFO algorithm, but the cell FIFO algorithm can take considerably longer during congested conditions.

The second enhancement is the modelling of explicit lanes and lane striping on the last cell of the approach. Lanes become blocked if the lane discharge capacity is exhausted and when a stream spills back onto the approach in the lanes with permission for that stream. Lanes remain open if the steam which spills back does not have permission. Two algorithms are provided for the relaxed FIFO discipline and for the cell FIFO discipline. The discharge capacity and the local supply are distributed according to lane capacities. The capacity of shared lanes is split evenly among the streams with permission, with some positive partial demand, and with some remaining downstream local supply. This distribution is less efficient, and often requires more iterations to meet the stopping criterion.

The diverge network was most realistically modelled with the relaxed cohort FIFO algorithm with explicit lanes, though it took nearly twice as long to simulate as the cohort FIFO algorithm without lanes. Part of this increase in simulation execution time was due to the increased level of congestion.

None of the intersection algorithms could capture the asymmetry of the lane drop at the merge or the effects of anticipatory lane changing upstream of the lane drop.

The realistic intersection representation improved the modelling of spillback on the oversaturated approaches. However, there is no mechanism in the cell-transmission model to capture the effects of gap acceptance for an opposed permissive movement. Therefore, the capacity reduction must be specified in the model input. Modelling explicit lanes permitted the specification of lane capacity reductions, which can be more restrictive to shared or exclusive left-turn lanes, for example. However, determining the temporal, lane-specific capacity reductions is not trivial, and the difficulty is compounded by the fact that CellNetLoad does capture spillover, which is dependent on the cell length. During heavy congestion the cell FIFO algorithms and the algorithms with explicit lanes can have very long execution times, far longer than even the microscopic simulation used as the benchmark.

Chapter 7

Simulation of a Commuter Corridor

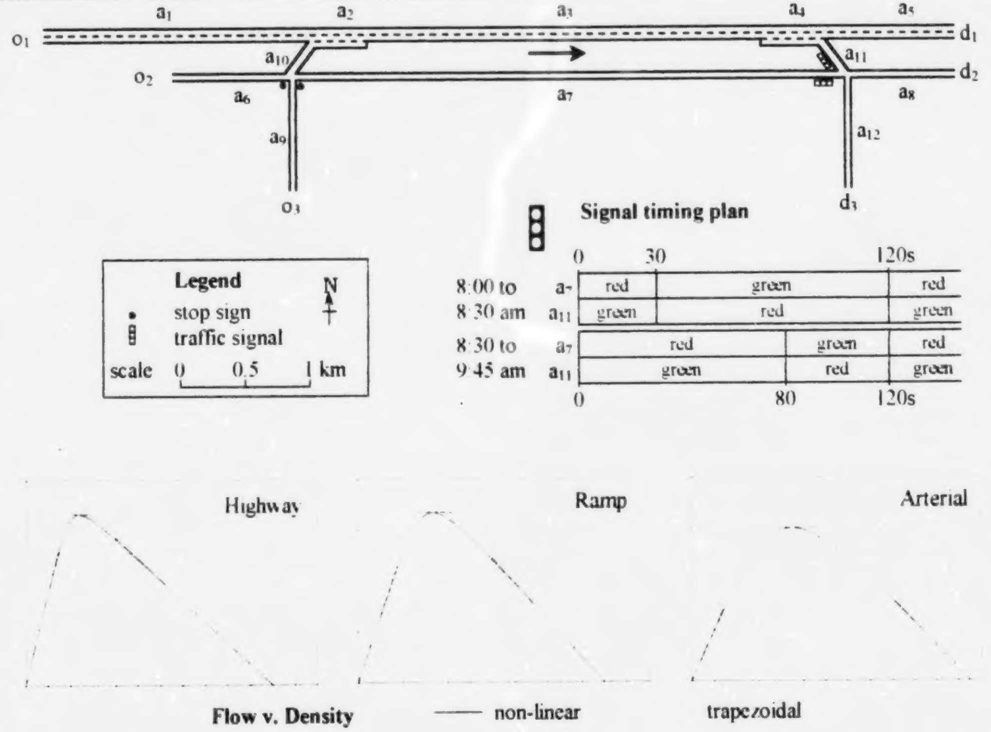
7.1 Introduction

The cell-transmission model was developed in order to approximate the LWR hydrodynamic model using a computer implementation. The potential for a parallel implementation implies that the temporal flow, density and travel time could be estimated for a very large network. The test networks, which were presented in previous chapters, were designed to isolate network features for easier analysis and comparison with a benchmark. In this chapter, the performance of the cell-transmission model is evaluated for a network with interactions among the congestion caused by separate network features.

A simple network with multiple origins and destinations was conceived in order to demonstrate the phenomena of queuing and spillback caused by a freeway merge bottleneck and by an oversaturated signalised intersection. This network and the demands are described in the following section. The simulation results are analysed and the simulation execution times are discussed.

7.2 Corridor network description

The corridor consists of an eastbound freeway and a parallel arterial (Figure 7.1). The 2-lane, eastbound freeway is 7 km long. An on-ramp merge is located at the 2-kilometre milepost with a 0.5-kilometre acceleration lane. An off-ramp is located at the 6 km milepost, with a 0.5-km deceleration lane. The single-lane, parallel arterial runs eastward for 850 metres until is meets a northbound arterial road and the on-ramp at a stop-controlled intersection. A signalised intersection provides alternating priority to the parallel arterial and the off-ramp.


a1
a2
a3
a4
a5
o1
d1
a10
a11
o2
d2
a6
a7
a8
a9
a12
o3
d3
Legend
N
stop sign
traffic signal
scale 0 0.5 1 km
Signal timing plan
0 30 120s
8:00 to 8:30 am
a7 red green red
a11 green red green
8:30 to 9:45 am
a7 red green red
a11 green red green
0 80 120s
Highway
Ramp
Arterial
Flow v. Density
non-linear
trapezoidal

Facility	v_f (km/h)	q_{max} (veh/h/lane)	v_c (km/h)	k_j (veh/km/lane)
	Speed-flow-density parameters			
Freeway	110	2000	85	110
Ramp	80	2000	60	110
Arterial	60	1800	40	125

Demand (vph) from 8:00 am to 9:40 am

	d_1	d_2	d_3
o_1	2400	450	450
o_2	300[1]	300[2]	
o_3	600[1]		

[1] ends at 8:20 am [2] follows arterial path

Figure 7.1 Corridor network

Traffic is loaded on a single set of paths between the origin-destination pairs beginning at 8:00 am. Congestion first occurs at the on-ramp merge as 900 vph merge with 3300 vph on the freeway. The 2-lane bottleneck has a capacity of 4000 vph at the end of the acceleration lane. A queue is expected to spill back on the freeway upstream of the on-ramp. At 8:20 am the on-ramp demand ends, and it is expected that the freeway will begin to recover.

Meanwhile, a short green time for the off-ramp results in oversaturation. A demand of 900 vph to the off-ramp receives only a 30-second green phase of a 120-second cycle length, creating a bottleneck with a capacity of 500 vph (2000 vph * 30 s / 120 s). At 8:30 am the signal timing is changed, such that the off-ramp receives an 80-second green phase, and the arterial receives only 40 seconds. The off-ramp now has capacity 1333 vph, allowing for the queue to dissipate.

The network was carefully configured to demonstrate the strengths and weaknesses of each model. An uncongested network would not expose the deficiencies of the more abstract models. A highly congested network, mainly due to vehicle-to-vehicle interactions, such as lane changing and permissive left-turns at intersections, would exaggerate the weaknesses of the flow-based model and the cell-transmission model.

7.3 Simulation results

A microscopic simulation program, called INTEGRATION, was chosen as the benchmark against which the accuracy of CellNetLoad is evaluated. The choice of a simulated benchmark over observations of existing traffic networks was motivated by the flexibility to customise the network and the traffic demand, as well as the ability to perform multiple runs with different random number seeds.

INTEGRATION uses car-following logic that is consistent with macroscopic speed-flow-density parameters. The speed of each vehicle is updated each deci-second based on the current headway of the vehicle, respecting the acceleration and deceleration constraints. Lane changing is modelled explicitly. These features imply a better approximation to reality than CellNetLoad which generally uses a much more coarse time discretisation, permits infinite acceleration and deceleration, and ignores the effects of lane changing.

The simulation results for each arc, including the inflow, outflow, density and travel time, are presented on a separate figure. The observations of the freeway arcs 1 through 4 are shown on Figure 7.2 through Figure 7.5. The observations of the on- and off-ramp are shown on Figure 7.6 and Figure 7.7. Each measure was averaged during two-minute intervals, which coincides with the cycle length of the signal. The maximum value of the inflow and outflow axes corresponds to the arc flow capacity. The maximum value of the density axis corresponds to the jam density. Figure 7.8 contains the temporal path travel times for the two paths joining origin 1 to destination 2 in one-minute intervals.

7.3.1 INTEGRATION simulation results

Congestion was caused by lane-changing manoeuvres. Lane-changing allows vehicles to briefly occupy two lanes, resulting in a capacity reduction. Mandatory lane changes occur within a window defined by a soft wall and a hard wall. A soft wall is the location in a lane at which a driver first realises that a lane change must be made in order to continue on the known path. The hard wall represents the location farthest downstream in the same lane at which the lane change manoeuvre must be made. A vehicle will slow down as it approaches the hard wall, blocking flow in that lane until it finds a suitable gap in the adjacent lane. Drivers become increasingly aggressive in accepting gaps in the adjacent lane as they approach the hard wall. Aggressive lane-changes can force the following vehicle to brake, thereby incurring some delay.

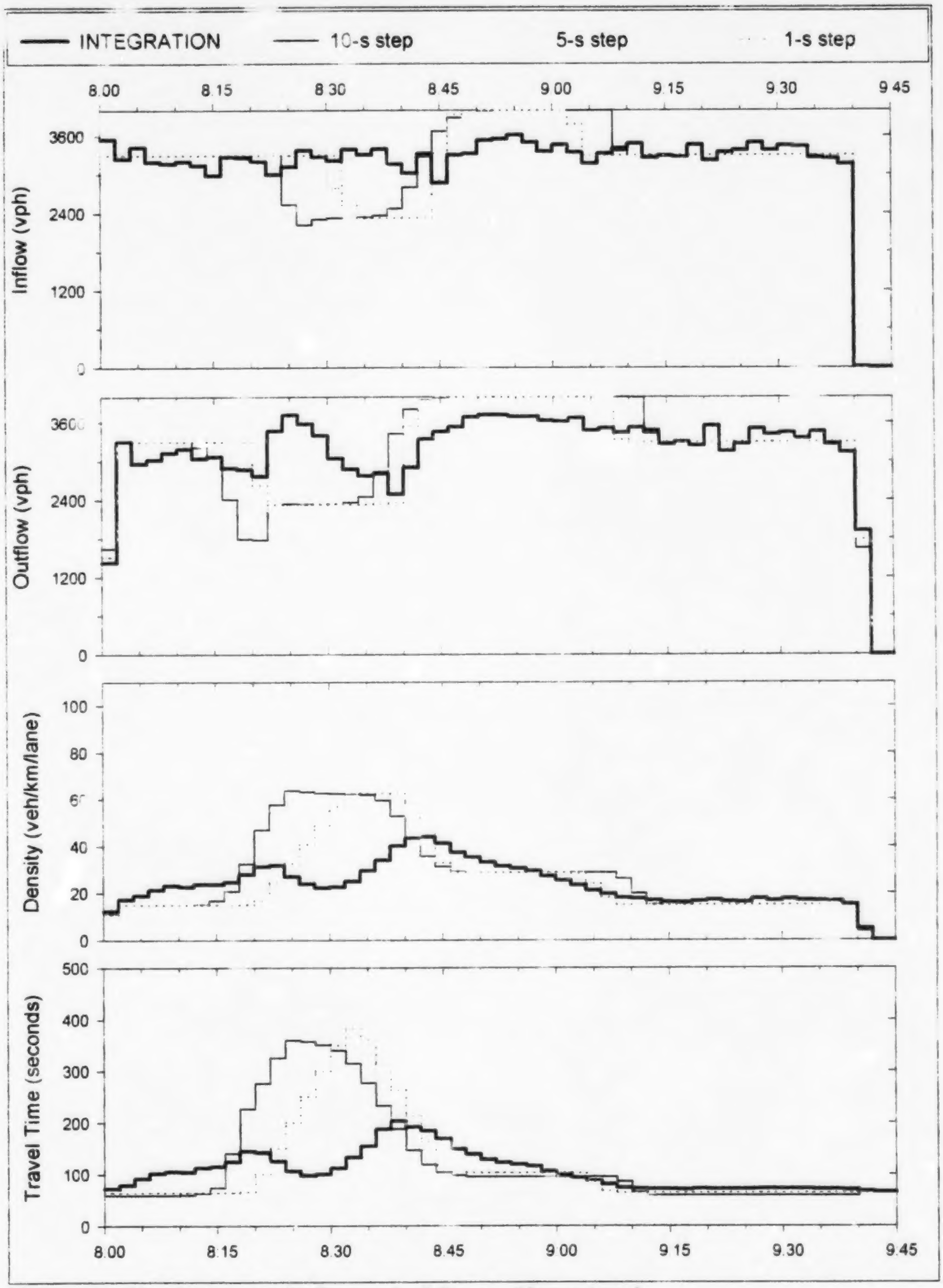


Figure 7.2 Simulation results for arc 1 of the corridor network

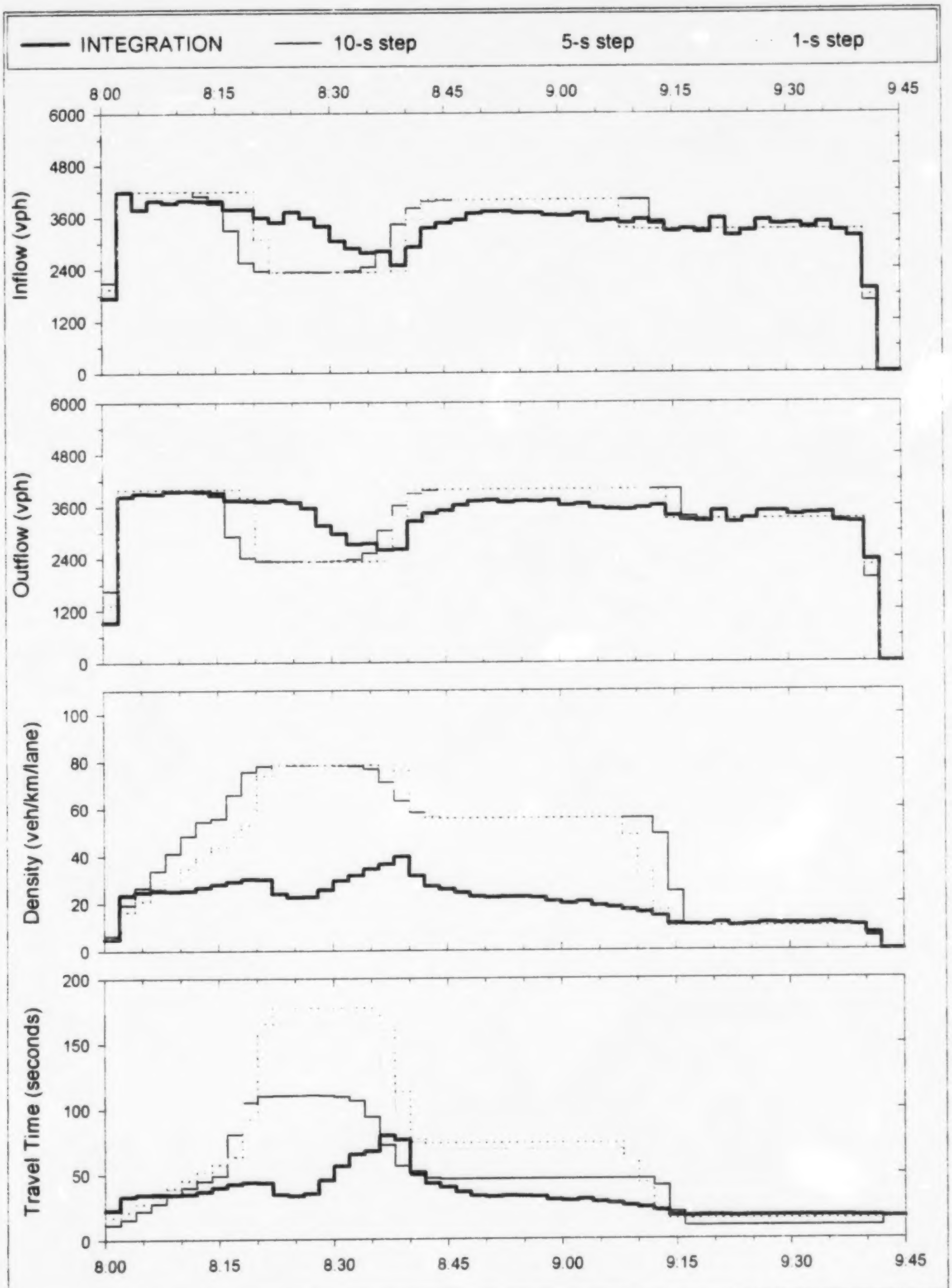


Figure 7.3 Simulation results for arc 2 of the corridor network

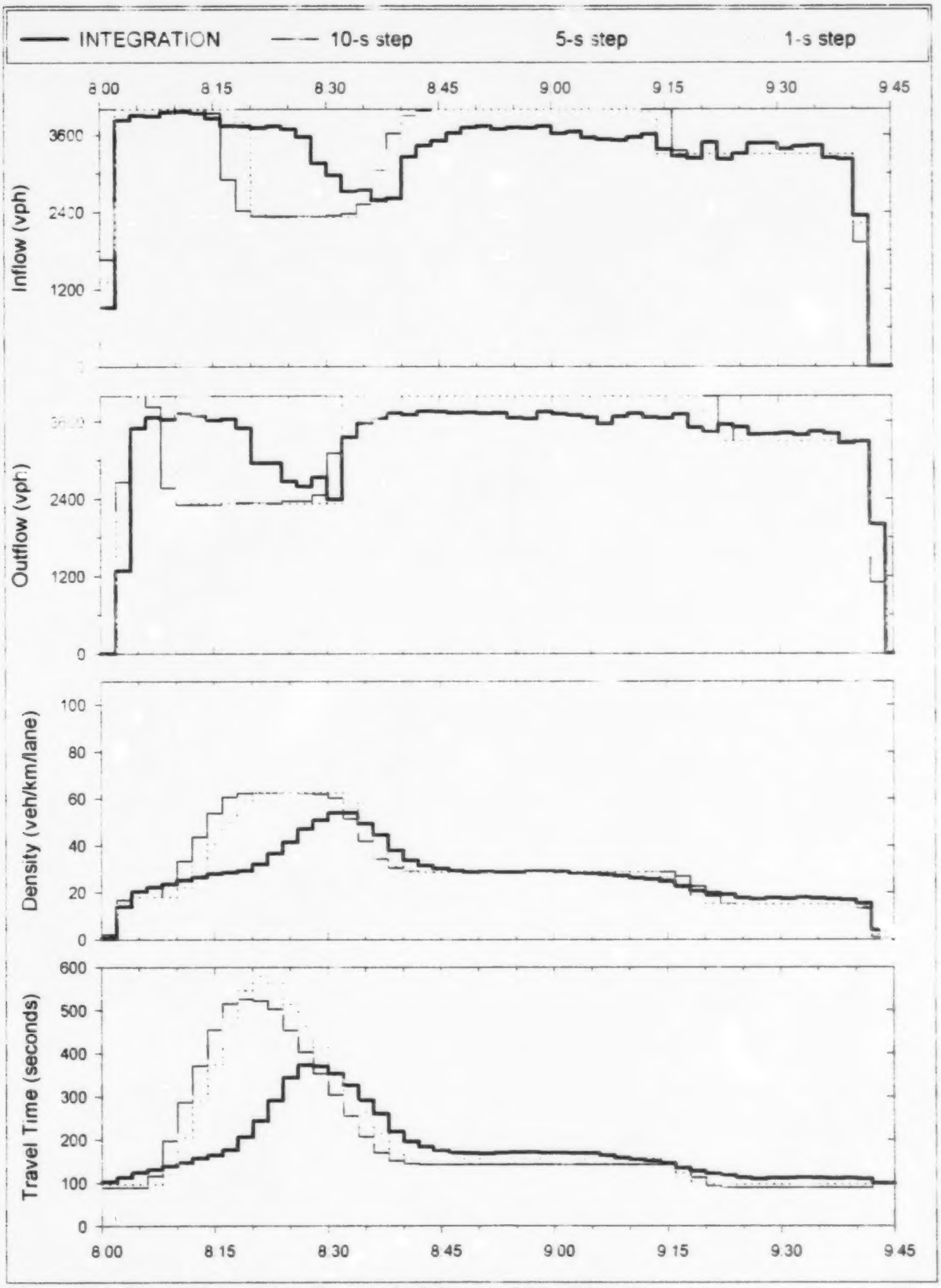


Figure 7.4 Simulation results for arc 3 of the corridor network

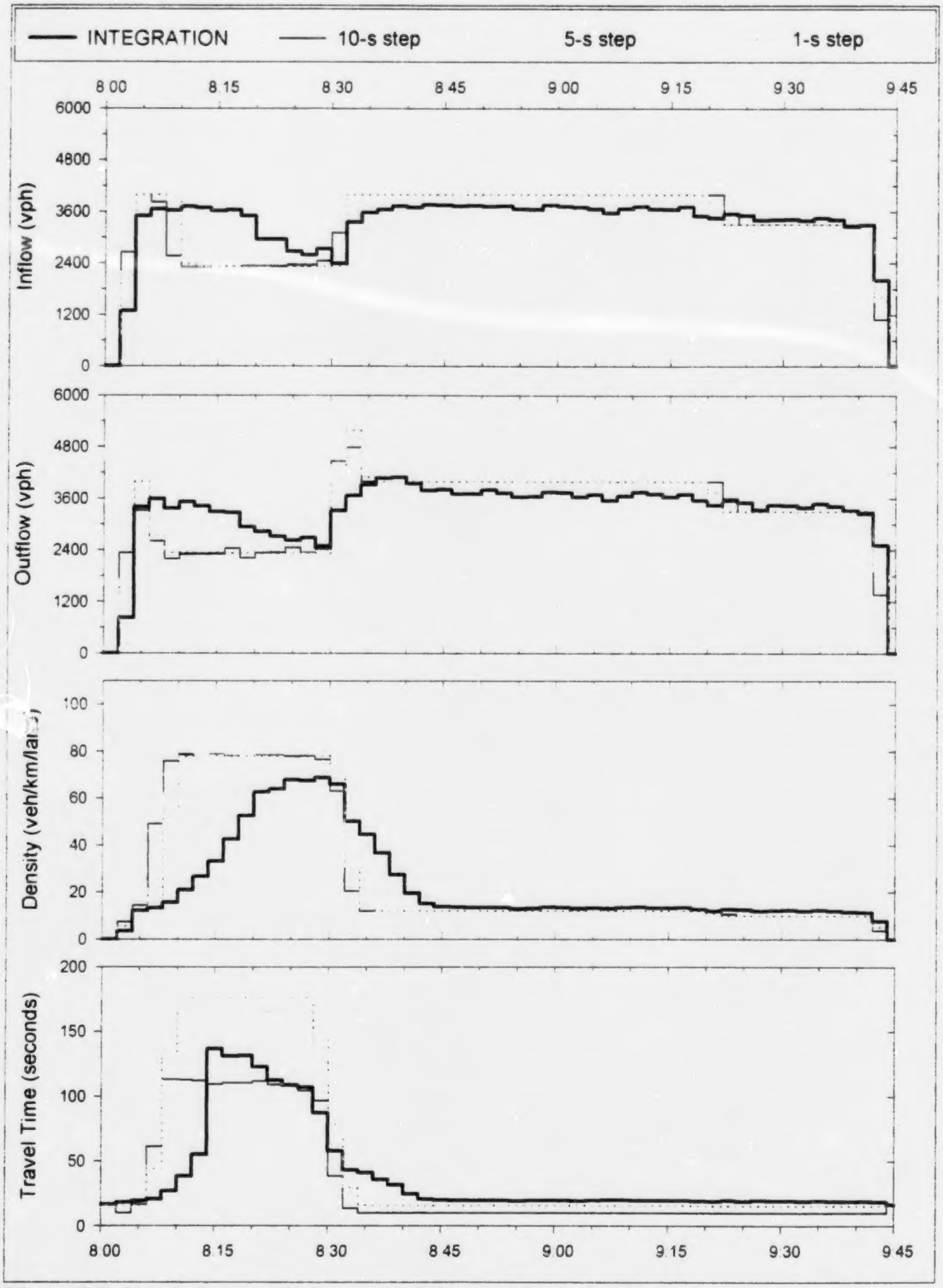


Figure 7.5 Simulation results for arc 4 of the corridor network

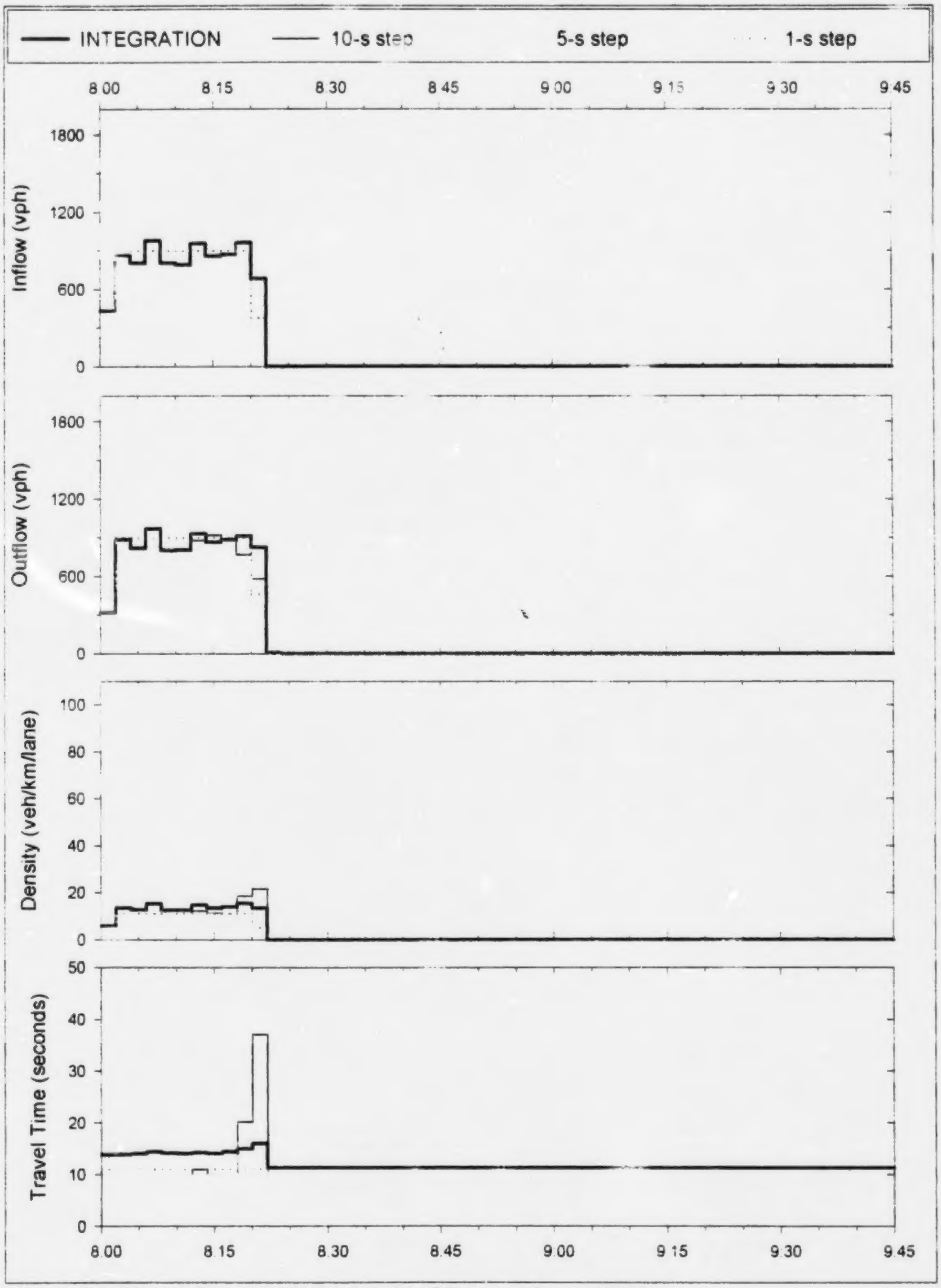


Figure 7.6 Simulation results for arc 10 of the corridor network

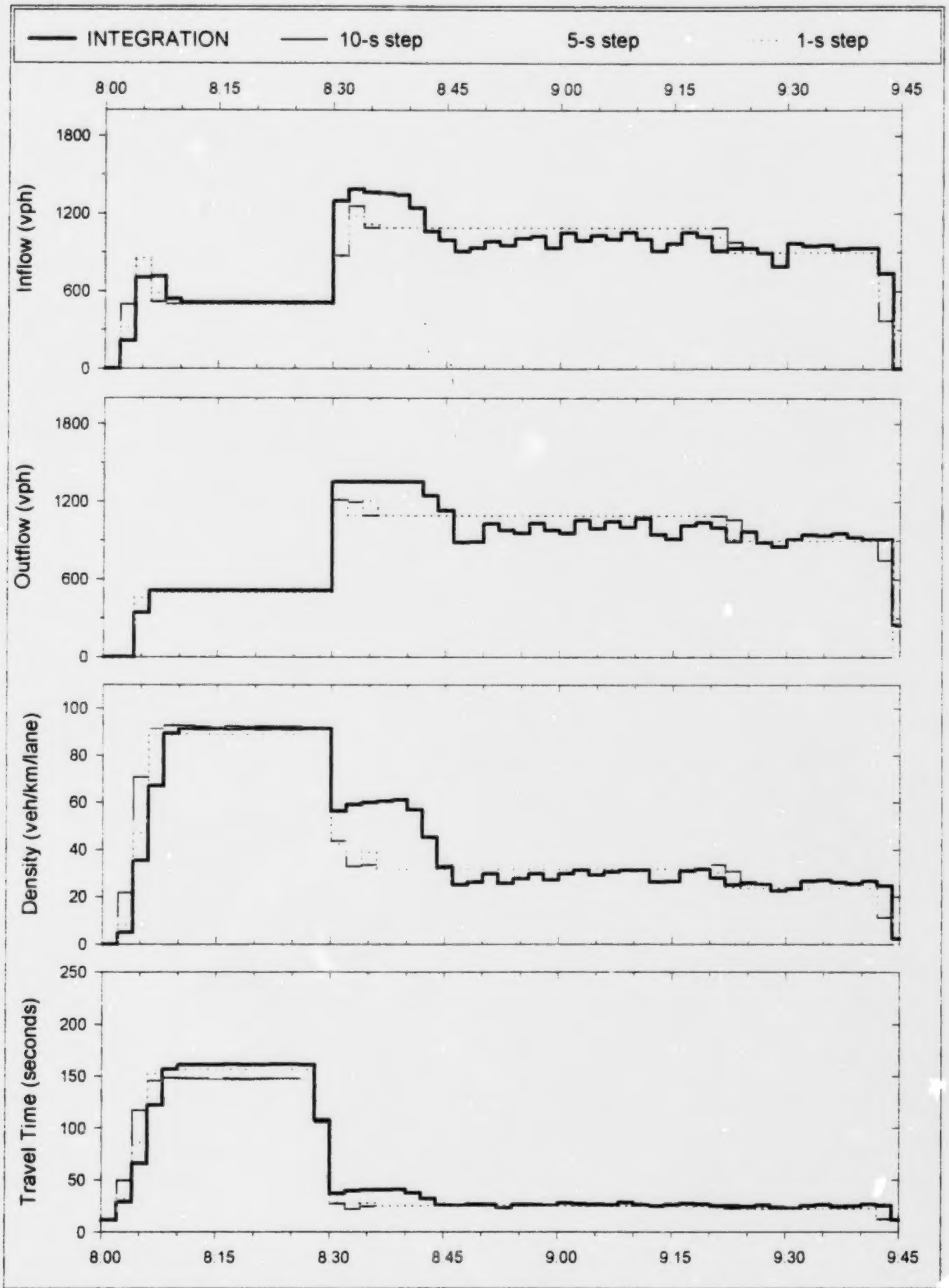


Figure 7.7 Simulation results for arc 11 of the corridor network

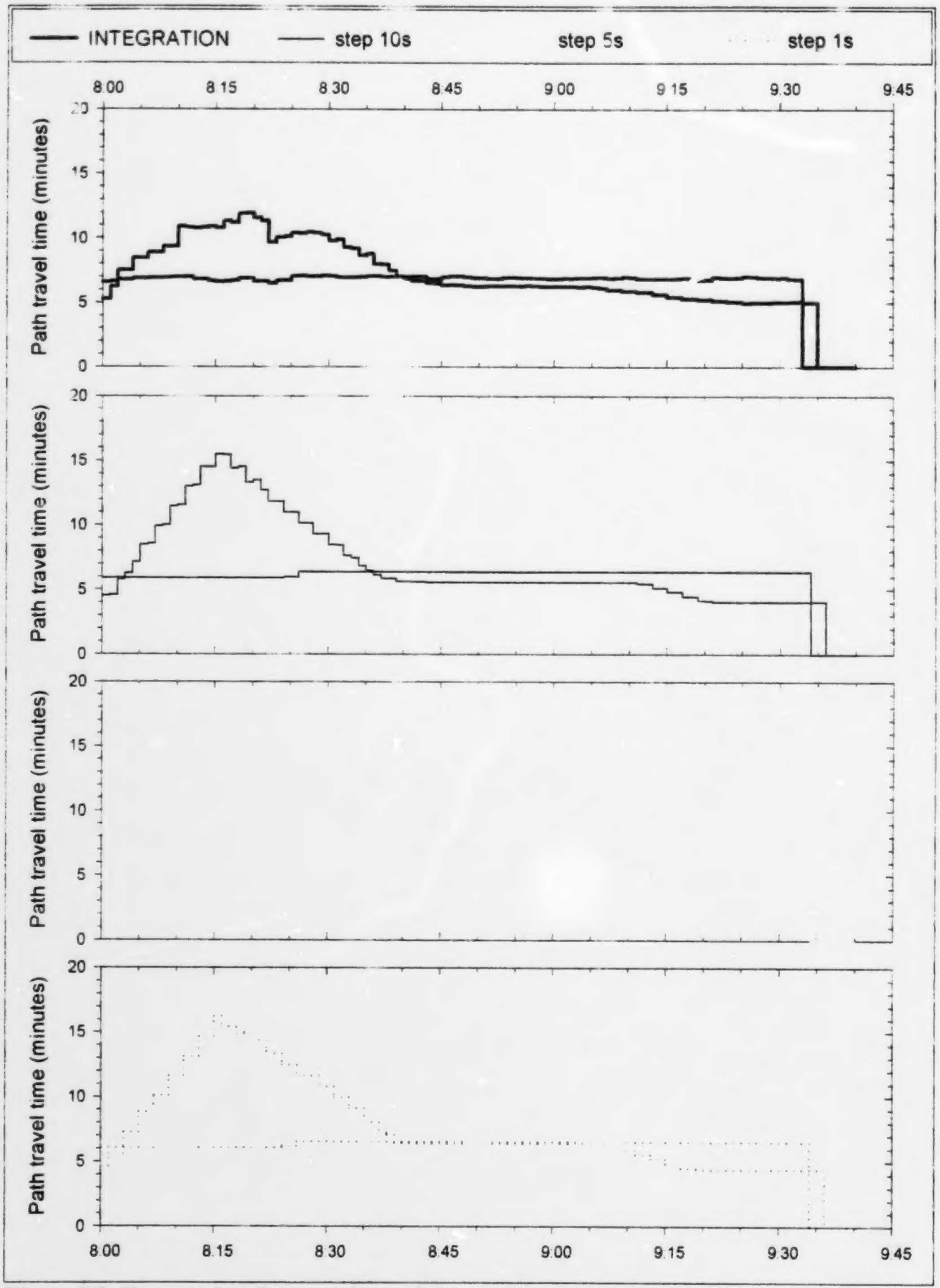


Figure 7.8 Temporal path travel times of the corridor network

Another source of congestion is flow control at intersections. Queues build during the red phase on a signalised approach to an intersection. If the demand exceeds the reduced capacity, then the queue fails to clear during the green phase. The queue length will grow each cycle and spill back on upstream arcs. Gap-acceptance logic at intersections results in delay and in a capacity reduction to approaches which contain an opposed, permissive movement.

The statistics were compiled over 2-minute collection intervals using arc entrance and exit times for each vehicle. The 2-minute collection interval was chosen to coincide with the cycle length of the signal to avoid oscillations in the flow which would obscure graphical comparisons. The average arc inflow (outflow) is simply the number of vehicles which enter (exit) the arc during the collection interval. The average arc density is calculated by summing the travel time spent by each vehicle on the arc during the collection interval, then dividing the sum by the collection interval length. The average arc travel time is calculated by summing the arc travel times of all vehicles which *entered* the arc during the collection period, then dividing the sum by the average arc inflow.

Several stochastic elements created "noise" in the data. Departure headways were exponentially distributed to provide a realistic distribution of gaps for proper gap-acceptance behaviour, which is itself a stochastic phenomenon depending on driver aggressiveness. Therefore, the observed arc inflow and arc outflow varied from one collection interval to the next, even if the mean flow rate was the same. The results of five simulation runs, each with a different random number seed, were averaged in order to filter out the "noise."

As shown on Figure 7.6 of arc 10, the merging flow takes priority, causing a queue to build up on the freeway. Merging vehicles on the acceleration lane of arc 2 aggressively force their way into the mainline freeway flow before the end of the acceleration lane. Mainline vehicles are forced to brake and propagate congestion upstream on arc 1 (Figure 7.2). Due to the priority of the on-ramp, only 92.5 percent of the freeway demand exits arc 1 (3050 vph outflow / 3300 vph demand).

The outflow of arc 3 (see Figure 7.4) never reaches the saturation flow rate due to mandatory lane changing of vehicles which exit the freeway. The off-ramp demand is equally distributed on both lanes of arc 1 at origin 1. Only 830 vph of the off-ramp demand, which enters as part of the freeway flow, reaches arc 3. Approximately half of the vehicles destined for the off-ramp must perform lane-changes on arc 3 from the left lane to the right lane. This reduces the outflow capacity of arc 3 from 4000 vph to 3675 vph.

Major congestion is caused by the restrictive signal timing at the off-ramp, which receives only 30 seconds of green time of a 120-second cycle length. Shockwave analysis reveals that a constant capacity flow of 500 vph inflow to the off-ramp would produce a cyclic queue 148 m in length, which would not spill back to the freeway. The actual initial inflow rate to the off-ramp is 784 vph of the 900-vph demand, since some of the demand accumulates on arcs 1, 2, and 3. This oversaturation of the off-ramp signal is sufficient to spill back on the deceleration lane of arc 4 during a portion of the third cycle, and for the entire duration of the sixth and subsequent cycles. Note that the density of arc 11 in Figure 7.7 remains peaked at 91 veh/km/lane, which is the jam density (110 veh/km/lane) during three quarters of a cycle and the density-at-capacity (33.3 veh/km/lane) during one quarter of the cycle. After the deceleration lane was filled, vehicles destined for the off-ramp spill over into adjacent lanes on the freeway, obstructing the through vehicles. The outflow from arc 4 (Figure 7.5) gradually decreases from nearly 3600 vph at 8:07 am to 2450 vph at 8:28 am, 500 vph of which exit the off-ramp.

This reduction in the outflow of arc 4 causes congestion to propagate upstream as a series of shockwaves. A congestion-forming shockwave propagates backward from the off-ramp at the start of each red phase, followed by a recovery fan as the platoon discharges during the green phase. The dissipation of the shockwave, due to deceleration constraints, and the spreading of the fan results in an aggregate shockwave which propagates backward. This aggregate shockwave causes a gradual decrease in the flow. The outflow of arc 3 (Figure 7.4) begins to decline at 8:18 am The aggregate shockwave reaches arcs 2 and 1 at 8:26 am and 8:30 am, respectively, as seen by the rising density on Figure 7.2 and Figure 7.3. The inflow to arc 1 drops at 8:44 am, indicating that the shockwave has reached the origin.

At 8:30 am the signal timing changes to favour the off-ramp with a green time of 80 seconds. The outflow from the off-ramp is first constrained by the capacity of 1333 vph, and later by the inflow of 900 vph plus the accumulated demand in the upstream congestion. The aggregate recovery shockwave propagates back like the congestion shockwave, first arriving at arc 3 at 8:32 am, next at arc 2 at 8:40 am, then at arc 1 at 8:42 am, and finally at the origin at 8:46 am. The flow climbs to capacity following the recovery shockwave until the queues have dissipated, and the flow reverts to the demand. This final forward-moving recovery shockwave starts at the origin at 8:56 am, and reaches the end of arcs 1, 2, and 3 at 9:14 am, 9:16 am, and 9:20 am, respectively. It can be shown with analytical calculations that the shockwave speed respects the Rankine-Huginot jump condition [2.13]. So shockwave propagation is consistent with the LWR hydrodynamic model.

The arc travel times are correlated to the density values, which can be seen by visual inspection of the figures. The arc travel times begin at the free travel time, then increase during the congestion and subside as recovery takes place. The travel-time peak begins earliest on arc 11, and subsequently later on the upstream freeway arcs. The duration of the peak is shortest for arcs 1 and 2, which recover soonest, and longer for arcs 3, 4, and 11. Also, note the presence of a second travel-time peak on arcs 1 and 2 due to the congestion caused by the merge flow.

An estimated path travel time can be derived recursively from the arc travel times. The path travel time is equal to the arrival time at the end of the first arc, given a trip departure time, plus the path time from the beginning of the next arc, with a departure time defined by the arrival at that arc. More intuitively, a vehicle probe starts a trip on a given route at a specified departure time and uses the current estimated arc travel time to traverse each arc as it arrives at the entrance to each arc.

The path travel times for two paths are shown in Figure 7.8. Note that the arterial path travel time (arcs 6, 7, and 8) is relatively constant, except for a brief dip due to the end of the on-ramp demand, which is then countered by the increased delay incurred at the downstream signal. The path which enters and exits the freeway (arcs 6, 10, 2, 3, 4, 11, and 8) has a characteristic peak due to the congestion. The specified temporal demand does not produce equilibrium flows. All demand from origin 2 to destination 2 takes the arterial route. Given the current temporal path demands, the best path choice would be the freeway route for a departure prior to 8:02 am or after 8:39 am The arterial route is faster between these two times. By moving a portion of the demand from the arterial route to the freeway route before and after the congestion, then obtaining the temporal path travel times, an equilibrium solution could be found iteratively.

7.3.2 CellNetLoad simulation results

CellNetLoad, an implementation of the cell-transmission model, generates simulation results for a series of time discretisations that are more coarse than INTEGRATION. Several enhancements to the cell-transmission model were available in CellNetLoad. These enhancements were activated so that the simulation results could be compared with the benchmark. The first section discusses the results of the basic cell-transmission model with particular emphasis on the effect of the time step. The following section chooses the most appropriate enhanced intersection algorithm based on the simulation results. The third section analyses the results for the simulation with enhancements to the flow propagation, which include a non-linear equilibrium flow-density relationship that is consistent with INTEGRATION, the accurate modelling of arc storage, and the discretisation of flow.

7.3.2.1 Effect of time discretisation

CellNetLoad was run first with no enhancements activated. The simulation was repeated with time discretised into steps of 10 s, 5 s, and 1 s. The simulation results are shown on Figure 7.2 through Figure 7.8, for comparison with INTEGRATION.

The cell length is defined as the distance travelled during one time step at the free speed. Since the arc is composed of an integer number of cells, the arc length is modelled with an error that is up to one cell length in size. The cell length for each facility, and the resulting arc lengths and errors are shown in Table 7.1.

The first congestion appeared upstream of the lane drop. The density of arc 2 grew at 200 vph, which is the upstream demand 4200 vph = 3300 vph + 900 vph minus the 4000-vph capacity of arc 3. The inflow to arc 2 began to decline at 8:12:20 for the simulation with a 10-s time step, when the density reached the congested regime at 53.3 veh/km/lane, which implies an equilibrium flow of 1397 vph per lane. The local supply was distributed during each iteration of the node model according to the size of the first cohorts in the last cells of arcs 1 and 10 until enough cohorts had been sent to exhaust the local demand. The outflow of arcs 1 and 10 were constrained, so congestion occurred on both.

The smaller time steps postponed the decrease in the inflow of arc 2. The decrease occurred 6 minutes later at 8:18:40 for the simulation with a 5-s time step, and nearly 8 minutes later for the simulation with a 1-s time step at 8:20:10. The finer time discretisation produces a better approximation to the solution of the LWR hydrodynamic model, in which the inflow reduction would occur when the backward-moving shockwave reached the upstream end of arc 2. This time can be determined analytically by first finding the shockwave speed using the Rankine-Huginot jump condition. The Rankine-Huginot jump condition from fluid dynamics (Richards, 1956) states that the shockwave speed must equal the difference in flow divided by the difference in density between the states immediately upstream and downstream of the shockwave. The shockwave began when the flow first reached the lane drop at the downstream end of arc 2 at 8:01:22, then moved upstream at 1.54 km/h ([1400-1333] vphpl / [12.7 – 55.9] veh/km/lane). The shockwave reached the upstream end of arc 2 at 8:20:49. The exact location of this shockwave within a cell is unknown in the cell-transmission model but, as the cells are made smaller by a finer discretisation, the location is known with more precision.

Table 7.1 Arc discretisation and error in arc length

Time step		10 s			5 s			1 s		
Facility	Free speed	Cell length (m)								
freeway (f)	30.6 m/s	306 = 30.6 m/s × 10 s			153			30.6		
ramp (r)	22.2 m/s	222			111			22.2		
arterial (a)	16.7 m/s	167			83.3			16.7		
Arc no.	Length	Number of cells, **modelled arc length (m)**, and *error in arc length (m)*								
1 (f)	2000 m	6	**1833**	*167*	13	**1986**	*14*	65	**1986**	*14*
2 (f)	500 m	1	**306**	*194*	3	**458**	*42*	16	**489**	*11*
3 (f)	3000 m	9	**2750**	*250*	19	**2902**	*98*	98	**2994**	*6*
4 (f)	500 m	1	**306**	*194*	3	**458**	*42*	16	**489**	*11*
5 (f)	1000 m	3	**917**	*83*	6	**917**	*83*	32	**978**	*22*
6 (a)	850 m	5	**833**	*17*	10	**833**	*17*	51	**850**	*0*
7 (a)	4300 m	25	**4167**	*133*	51	**4250**	*50*	258	**4300**	*0*
8 (a)	850 m	5	**833**	*17*	10	**833**	*17*	51	**850**	*0*
9 (a)	850 m	5	**833**	*17*	10	**833**	*17*	51	**850**	*0*
10 (r)	250 m	1	**222**	*28*	2	**222**	*28*	11	**244**	*6*
11 (r)	250 m	1	**222**	*28*	2	**222**	*28*	11	**244**	*6*
12 (a)	850 m	5	**833**	*17*	10	**833**	*17*	51	**850**	*0*

A second bottleneck is located at the exit of the off-ramp due to the signal timing, which favours the arterial until 8:30 am. The oversaturated signal at the off-ramp produced a queue which first spilled back to arc 4 at 8:05:20 during the simulation with the 10-s time step. The shockwave arrived at arc 3 at 8:07:30. The spillback merged with the queue on arc 2 at 8:15:30, and forced the queues on arcs 1 and 10 to grow faster, even spilling back to origin 1 at 8:24:10. The dissipation of the shockwave, due to numerical error, and the spreading of the recovery fan during each green phase blend together into an aggregate shockwave with an observed speed of 20.6 km/h (2750 m in 8 minutes) on arc 3. This is consistent with shockwave analysis. The downstream state was the observed outflow from arc 3 of 1165 vphpl with a corresponding equilibrium density of 62.2 veh/km/lane. The upstream flow-state was the observed inflow to arc 3 of 2000 vphpl with a corresponding equilibrium density of 18.2 veh/km/lane. The predicted shockwave speed is 19.0 km/h in the upstream direction. The difference between the observed and predicted shockwave speed is attributed to numerical error which causes the shockwave to spread. The observed shockwave speed for the simulations with the 5-s and 1-s time steps were 20.1 and 20.0 km/h, respectively. The smaller time step decreases the amount of shockwave dissipation.

At 8:30 am the signal timing changed to favour the off-ramp. The off-ramp was nearly full with a stopped queue. The queue discharged during at the start of the green phase at the saturation flow rate of 2000 vph, creating some space for inflow to the off-ramp. The inflow is determined by the intersection algorithm as the minimum of the local supply and a portion of the discharge capacity of arc 4. At first, the inflow was constrained by the local supply, but the discharge capacity became the active constraint before the end of each cycle. Consequently, the outflow of the off-ramp did not remain at the saturation flow rate during the entire green phase. The restrictiveness of the diverge logic created a lower net outflow of 1090 vph from the off-ramp, compared to INTEGRATION which had an outflow of 1333 vph.

A backward-moving recovery shockwave arrived at arc 3 at 8:31:10, then at arc 2 at 8:46:20, next at arc 1 at 8:47:30, and finally at origin 1 at 8:53:40. This particular shockwave was an approximation of a fan, since the density decreased across the shock in the direction of flow. The approximation of the fan is a result of the piecewise-linear function used for the equilibrium flow-density relationship.

As soon as the backward-moving recovery shockwave met the back of the congestion, a forward-moving recovery shockwave formed. This shockwave, with increasing density across the shock in the direction of flow, exited origin 1 at 9:08:10, then arrived at arc 2 at 9:12:30, next at arc 3 at 9:16:00, at arc 4 at 9:22:40, and finally at arc 5 at 9:22:50. The observed propagation speed of 24.8 km/h across arc 3 is consistent with the predicted shockwave speed of 25.2 km/h ([2000-1650] vphpl / [28.9-15.0] veh/km/lane), considering that the observed times have a resolution of 10 seconds. The observed shockwave speed for the simulations with the 5-s and 1-s time steps were 25.5 and 25.1 km/h, respectively.

Several qualitative comparisons can be drawn with INTEGRATION. The congestion propagates faster, and recovers more quickly, but is more severe. Ignoring the capacity reduction due to lane changes forces congestion to propagate upstream more quickly, and recover more quickly. The merge logic creates a small queue on the on-ramp, which was not found in INTEGRATION. It also forces all of the lane-changing to occur at the lane drop without a penalty, which means that the congestion begins to form more slowly and further downstream, reducing the impact of spillback at the merge.

The FIFO condition is too restrictive for the mainline flow of the freeway at the off-ramp. This resulted in a faster propagation of the congestion upstream of the diverge. The increase in travel time on arc 1 due to the merge was obscured by the quick spillback from the off-ramp. Thus, only one large, early peak is visible for the travel time of arc 1, instead of the double peak found to occur in INTEGRATION.

The freeway path travel time is more peaked due to the restrictiveness of FIFO discipline at the diverge. The arterial path travel time is mildly underestimated for several reasons: the vehicles travel at the free speed during uncongested conditions due to the linear flow-density relationship; vehicles do not stop at the stop sign; no FIFO discipline is enforced for the stop-controlled intersection; and vehicles are permitted unbounded acceleration when leaving the queue at the signal. The arterial path is shorter than the freeway path from 8:04 am to 8:36 am, which is a shorter interval than predicted by INTEGRATION.

The finer space-time discretisation increases the number of time steps and decreases the length of the cells. Shockwave spreading is reduced, so the simulation results converge to the solution of the LWR hydrodynamic model in which shockwaves remain density discontinuities. Another impact is the reduced error in the modelled arc length. This improves the estimates of arc travel times, particularly for short arcs where the errors can be relatively large. Spillback is also modelled more accurately. A third impact occurs at intersections, where overtaking is permitted among packets in the same cohort. Smaller time steps produce smaller cohorts, so FIFO is more strictly enforced. This third impact was not particularly noticeable in the simulations of the corridor network.

Overall, the finer discretisation seems to increase the discrepancy between the path travel time estimates of CellNetLoad and INTEGRATION. This is explained by the fact that shockwaves spread in INTEGRATION due to acceleration and deceleration constraints. The reduction in spreading as the discretisation is made finer pushes the simulation results of CellNetLoad closer to the solution of the LWR hydrodynamic model and away from the INTEGRATION simulation results for this particular network. Because the mechanisms for the spreading of shockwaves are different, this finding cannot be generalised for other networks.

7.3.2.2 Effect of the enhanced intersection algorithms

The generalised intersection algorithm of CellNetLoad determines the flow across a node with multiple arcs in the backward star (a merge), in the forward star (a diverge), or both. The basic cell-transmission model does not permit a diverge cell to send flow directly to a merge cell, which forces intersections to be represented like interchanges, with a diverge on each approach separating the streams. This representation requires some logic to share the approach flow capacity and storage capacity among the streams as the partial demands change.

The first set of enhanced intersection models relaxes the cohort FIFO discipline, though FIFO discipline is strictly respected for each stream. The cell FIFO algorithm permits overtaking among packets in the same cell, even if they are in different cohorts. The relaxed cohort FIFO discipline permits overtaking only among packets that are at the front of the cohort queue, representing the local demand. Spillover affects all streams of an approach as soon as one steam spills back to an approach. The discharge capacity and the local supply are distributed to the competing streams in proportion to the partial demand of each stream.

The performance of the algorithms is evaluated in terms of the resulting partial flows at the intersections, although these flows directly affect the speed of the shockwaves that divide the flow states on the entire network. These sights include the on-ramp merge with the freeway, the off-ramp diverge from the freeway, the stop-controlled intersection, and the signalised intersection.

The partial flows at the stop-controlled intersection do not change with different intersection algorithms because there is no congestion. The signalised intersection is also unaffected by the intersection algorithm because the off-ramp approach and the arterial approach operate independently, and the local supply of the arcs in the forward star was not an active constraint. Likewise, the merge did not operate differently because the flow on the freeway and the on-ramp were each treated as a single traffic stream since both send flow exclusively to arc 2.

The diverge operates differently for the alternative intersection algorithms. The inflow to and the outflow from the off-ramp observed in the simulation with the cell FIFO algorithm matched INTEGRATION closely, as shown in Figure 7.9. The discharge capacity of arc 4 is distributed in proportion to the partial demands for the entire occupancy of cell 4, instead of the occupancy in the front cohort only. The spillover effect on arc 4 is also reduced, which is consistent with the findings in INTEGRATION. The relaxed cohort FIFO algorithm had a similar, but short-lived, effect. The relaxation of cohort FIFO affected the operation of the diverge only until the density of arc 4 exceeded the local demand.

The second set of enhanced intersection models have explicit lanes. Lanes are blocked if the lane discharge capacity has been exhausted, or if some partial demand is headed to an arc in the forward star with inadequate local supply and the lane striping gives permission to send flow from the lane to that arc. The discharge capacity and the local supply are distributed to the competing streams in proportion to the capacity of lanes with permission for each stream. Shared lanes are divided equally among all streams with permission on that lane and with some positive partial demand. FIFO discipline is respected in each lane and for each stream. The lane cell FIFO algorithm and the lane relaxed cohort FIFO algorithm correspond to the same relaxations of the cohort FIFO described above.

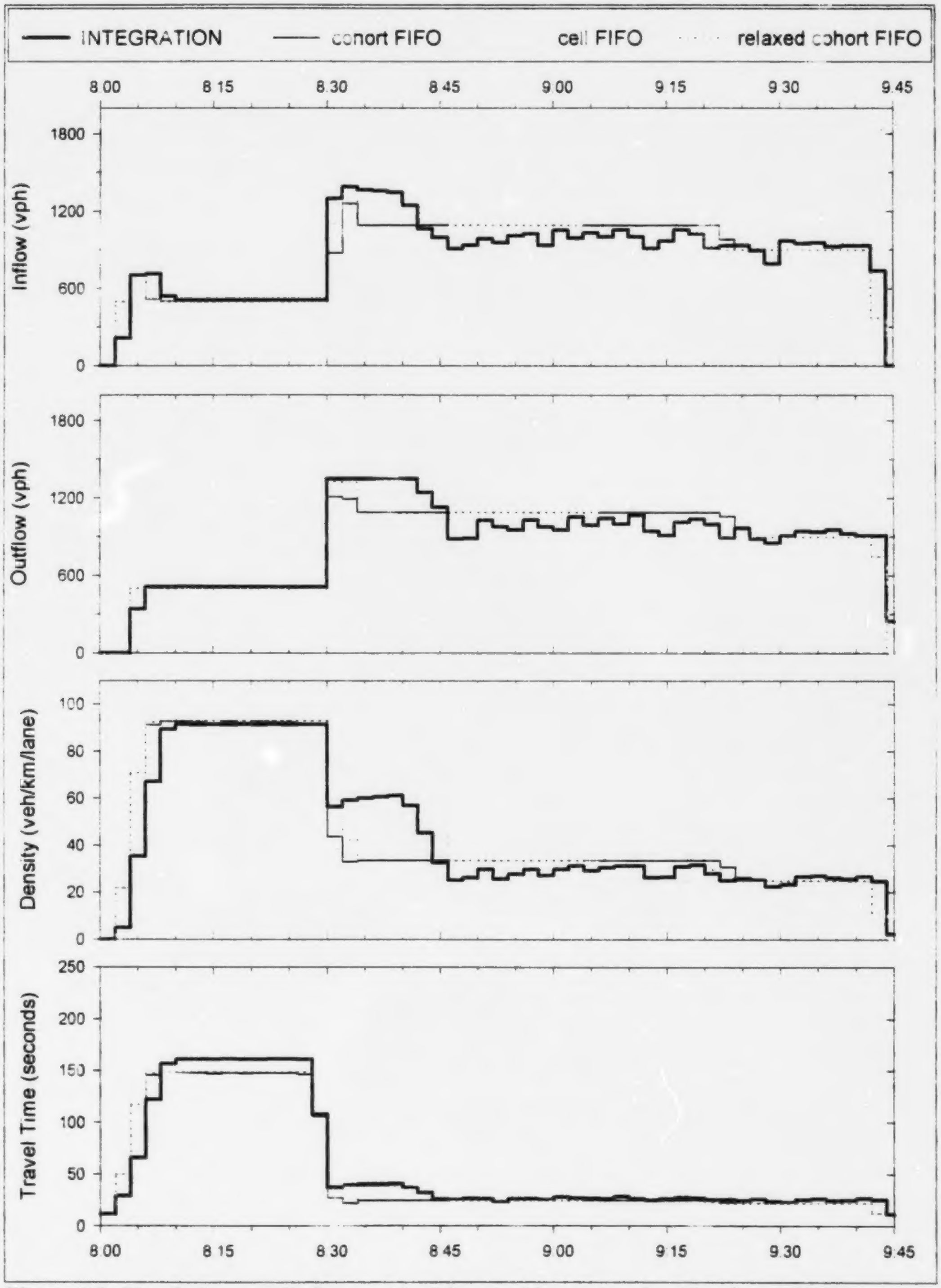


Figure 7.9 Simulation results for arc 11 of the corridor network with relaxed FIFO discipline

The stop-controlled intersection is unaffected by this change in the intersection algorithm because each approach is a single lane and the two streams on the eastbound approach have equal demands. The signalised intersection is unaffected for the same reasons.

There is no appreciable difference in the simulation results for the merge and diverge intersections when modelled with explicit lanes. The only difference occurred at the beginning of spillback. The spillback from the off-ramp was restricted to the deceleration lane on the freeway until the occupancy of arc 4 exceeded the local demand and spillover affected the mainline flow. No spillback occurred on the on-ramp. The on-ramp received a larger portion of the local supply when lanes were modelled explicitly because the partial demands consider only the first cohort in the cohort queue during each iteration of the intersection algorithm. The on-ramp demand ended before spillback would have eventually occurred. The differences in the simulation results are brief and differ by less than 2 percent.

When compared with the benchmark simulation results, it was found that some relaxation of cohort FIFO discipline improves the accuracy of the observed flows in CellNetLoad at intersections with multi-lane approaches. The simulation results obtained with the cell FIFO algorithm were found to be the closest match with INTEGRATION. The time step used in this analysis was 10 s. Smaller time steps would reduce the effect of the relaxation of FIFO discipline, eventually converging to the solution of the LWR hydrodynamic model with strict FIFO discipline for all flow. Obviously, the LWR hydrodynamic model predicts a different solution to the dynamic network loading problem compared to INTEGRATION, in particular for networks which include multi-lane approaches on which a relaxation of FIFO discipline is realistic.

7.3.2.3 Effect of the enhanced flow propagation

Several enhancements to the flow propagation of the cell-transmission model were implemented in CellNetLoad. Firstly, a non-linear equilibrium flow-density relationship, which is consistent with the speed-headway relationship used by INTEGRATION for the car-following logic, is available. Secondly, the storage capacity of the last cell in each arc can be adjusted to compensate for the error in modelling the arc length with an integer number of cells. Finally, flow can be modelled as discrete vehicles with probabilistic rounding to calculate flow. These enhancements are described in detail in chapter 5.

These enhancements were activated for the CellNetLoad simulation of the corridor network with the cell FIFO algorithm for intersections and explicit lanes. The non-linear equilibrium flow-density relationship was activated in combination with the modelling of exact arc storage. Another simulation was performed with the same settings, except that the enhancement for discretised flow was made active.

The impact of these enhancements is subtle but significant. Arc 4 is able to store 165 vehicles (0.5 km * 3 lanes * 110 veh/km/lane), instead of just 101 vehicles when the arc was modelled as only 306 m long. Overtaking is permitted for any packets within the single cell which represents arc 4. Because arc 4 could accommodate more vehicles, the through traffic is able to overtake a larger queue of vehicles queued for the off-ramp. In other words, spillover of the off-ramp is delayed, so the mainline flow past the off-ramp is congested later in the simulation. This reduces the speed of the shockwave which spills back to arc 3. In fact, the spill back never reaches the upstream end of arc 3, as shown by the solid grey line on the inflow diagram of Figure 7.10. The reduced congestion found in CellNetLoad suggests that the relaxation of cohort FIFO discipline permits too much overtaking and underestimates the effect of spillback in this particular network, compared to the benchmark.

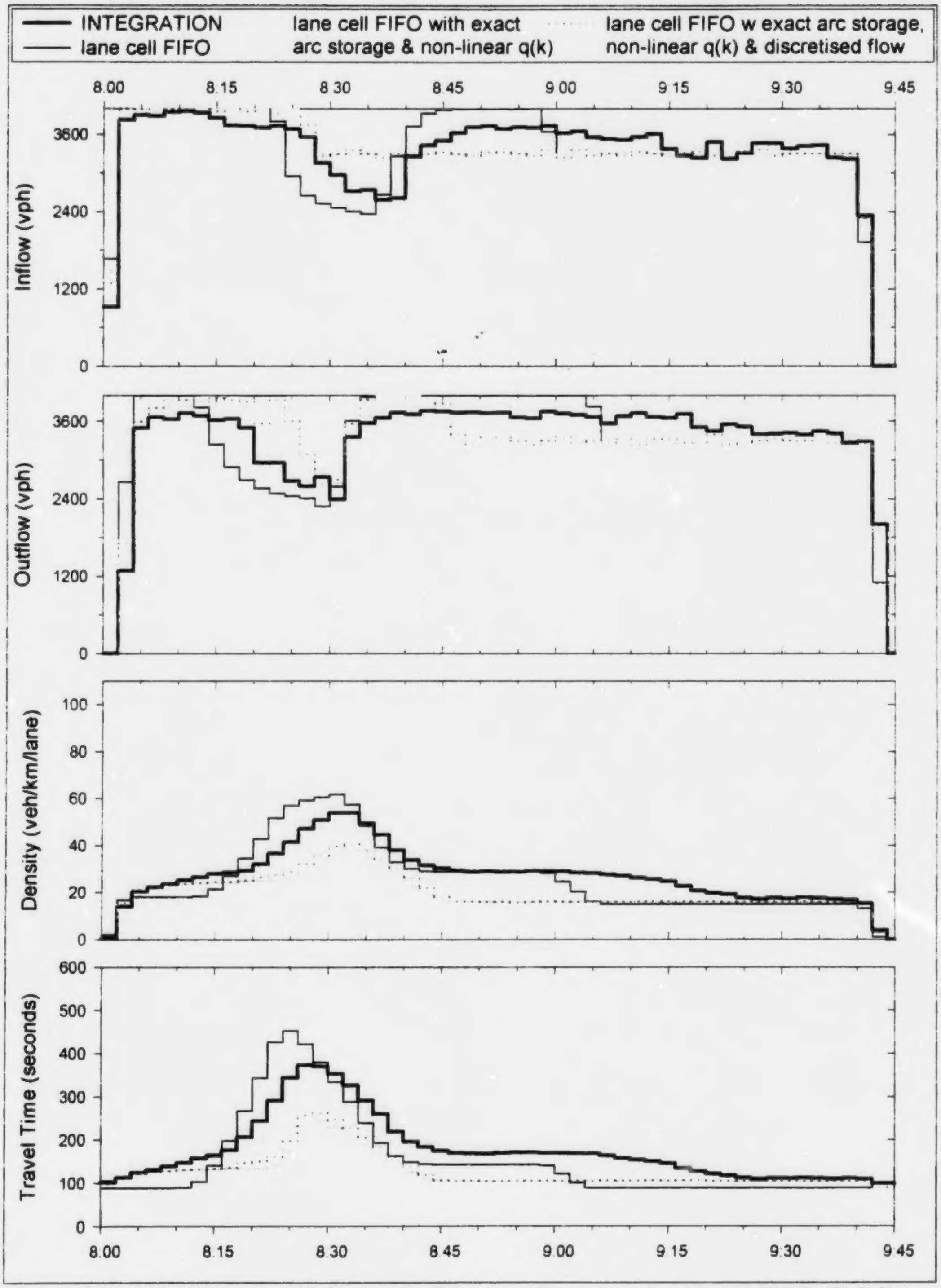


Figure 7.10 Simulation results for arc 3 of the corridor network with cell FIFO discipline and enhanced flow propagation

Another effect of the enhanced flow propagation, which is more pervasive, is that speeds are reduced in the uncongested regime due to the non-linear equilibrium flow-density relationship. Evidence appears on the travel time of arc 3 in Figure 7.10, which matches the travel time found by INTEGRATION at the beginning of the simulation before much congestion is present.

The combined effect of the reduction of the congestion and the increase in travel time during uncongested conditions can be found in the temporal path travel times (Figure 7.11). Although the magnitude of the temporal path travel times is smaller, the shape matches the benchmark much more closely for the simulation with the enhanced flow propagation than without. The underestimation of the congestion at the off-ramp affects the simulation results of arc 1 (Figure 7.12), which is not included in the evaluation of the temporal path travel times.

The activation of discrete flow in CellNetLoad adds stochasticity to the simulation due to the probabilistic rounding of flows. Though this implementation of discrete flow tends to restrict the flow due to a bias in the model, the effects if the flow restriction do not have a severe effect at this time step. The simulation results are shown as the dotted line in Figure 7.10, Figure 7.11, and Figure 7.12. Finer discretisations would produce greater flow restrictions. Hence, CellNetLoad does not converge to the solution of the LWR hydrodynamic model if the flow is discretised.

The contrast between the fit found in Figure 7.11 and the poor results in Figure 7.12 is a demonstration of an incorrect conclusion, which can be derived when only a portion of the simulation results are considered. This is a frequent concern when calibrating any model of network traffic flows, but it is also an obvious concern when validating a model.

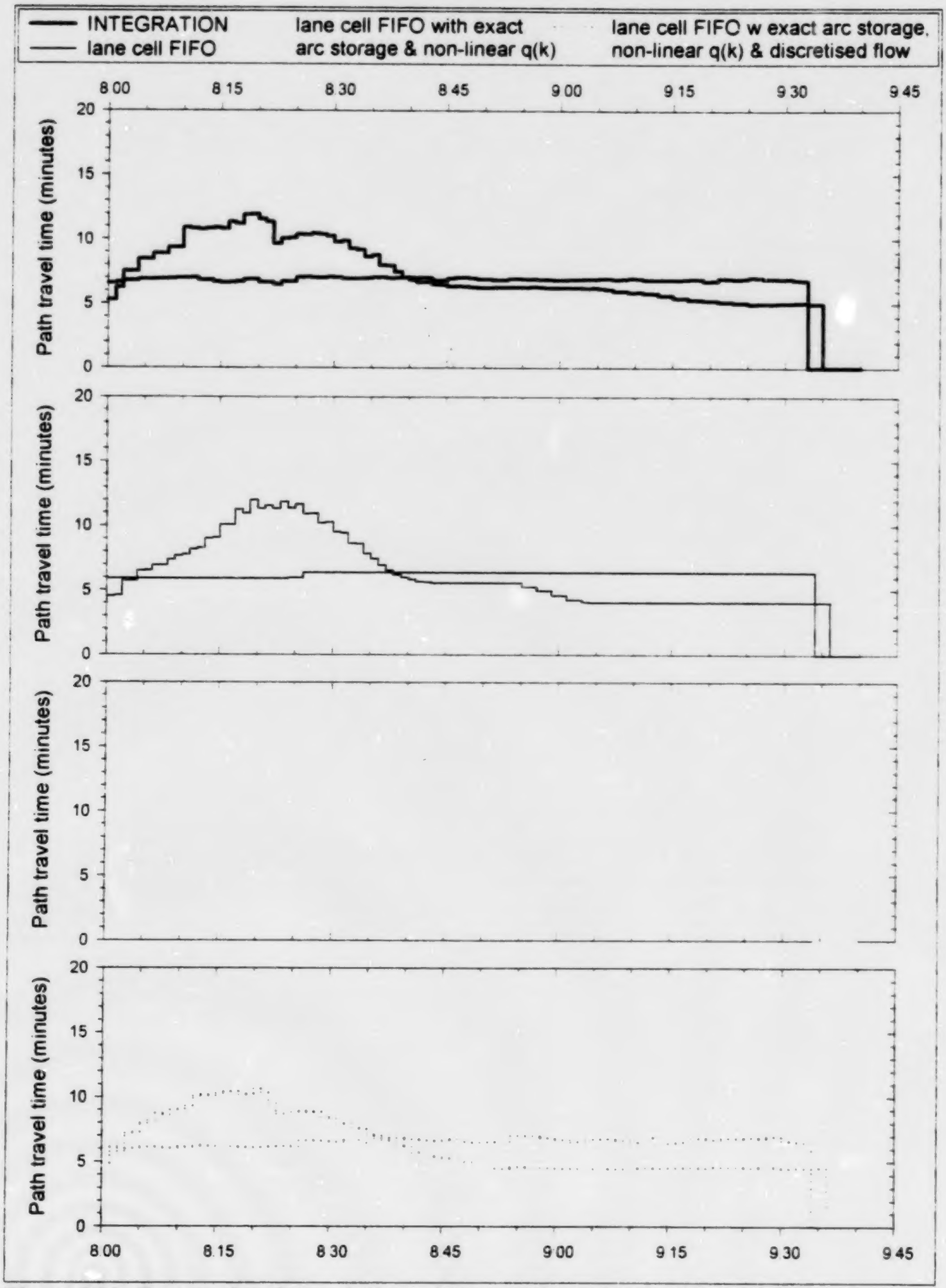


Figure 7.11 Temporal path travel times of the corridor network with cell FIFO and enhanced flow propagation

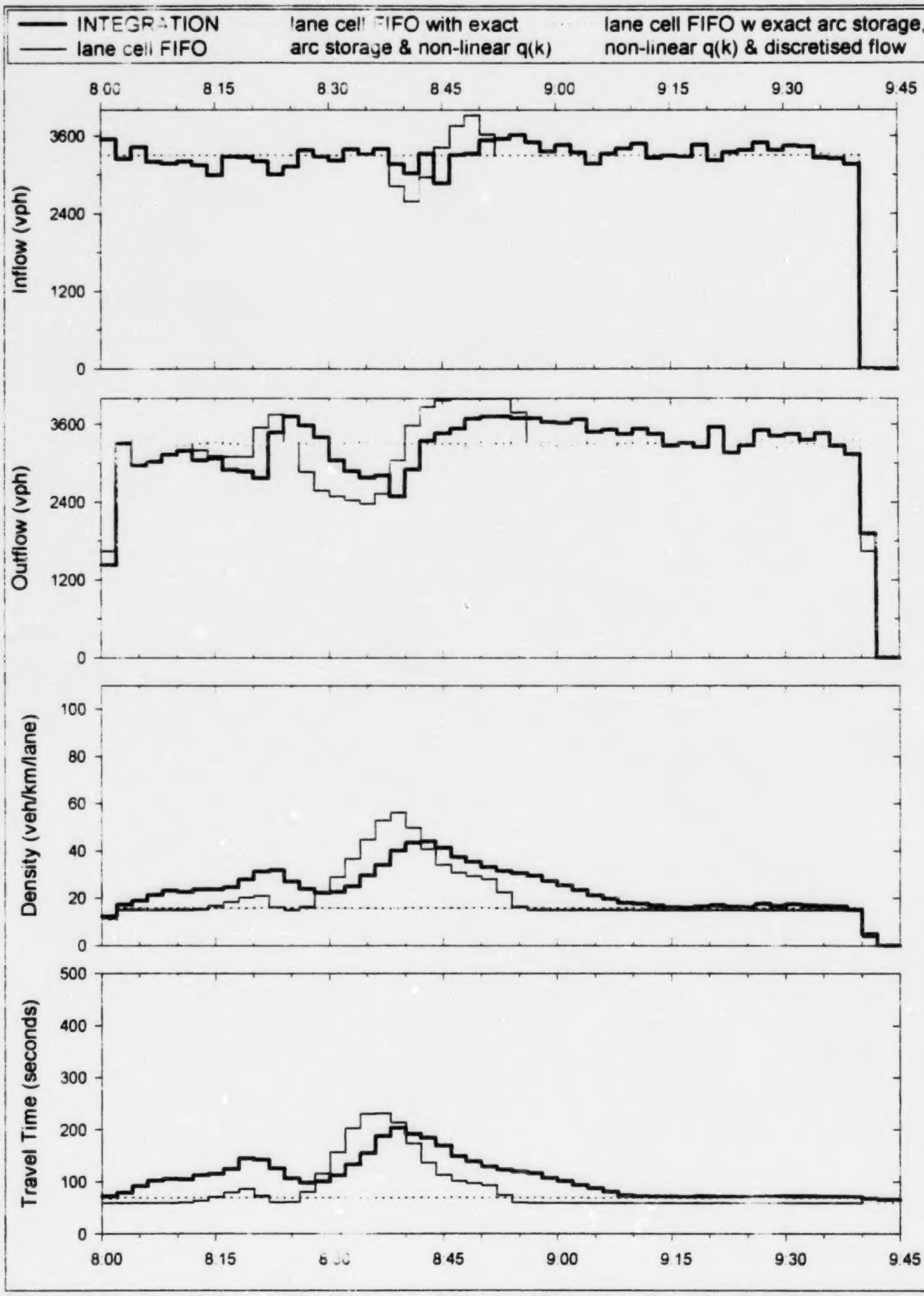


Figure 7.12 Simulation results for arc 1 of the corridor network with cell FIFO discipline and enhanced flow propagation

7.4 Simulation execution times

It was found from complexity analysis that the simulation execution time increased linearly with the number of cells in the network, the number of arcs in the network, and the number of output intervals, and it increased with the square of the number of time steps. The predicted simulation execution time and the observed simulation execution times for CellNetLoad are shown in Figure 7.13. The difference between the prediction and the observations is attributed to the intersection algorithm. The fact that some finely discretised simulations were shorter than the prediction is explained by the assumption that all cells are last cells. There is actually a savings in the execution time as the proportion of regular cells increases.

The simulation execution time of the cell FIFO algorithm (unfilled squares) is much larger than the cohort FIFO algorithm (unfilled circles). The cohort FIFO algorithm considers each cohort as a separate partial demand. So the algorithm iterates until the calculated flow from each cohort equals zero, until the local demand or the local supply are exhausted. The number of cohorts in the cell which are considered depends at which point the algorithm stops. The cell FIFO algorithm determines the partial demands once for all the cohorts in the cell, then iterates until the local demand or the local supply is exhausted. Although the algorithm iterates once per time step, the determination of the partial demands forces the algorithm to consider every cohort in the cell. During congested conditions, such as were found in the simulation of the corridor network, the cell FIFO algorithm is much slower.

The relaxed cohort FIFO algorithm (unfilled triangles) operates the same as the cell FIFO algorithm. It determines the partial demands once at the beginning of the algorithm, then iterates to find the partial flows. However, the determination of the partial demand only considers enough cohorts until the local demand is exhausted. So, the simulation execution time is closer to the cohort FIFO algorithm than the cell FIFO algorithm.

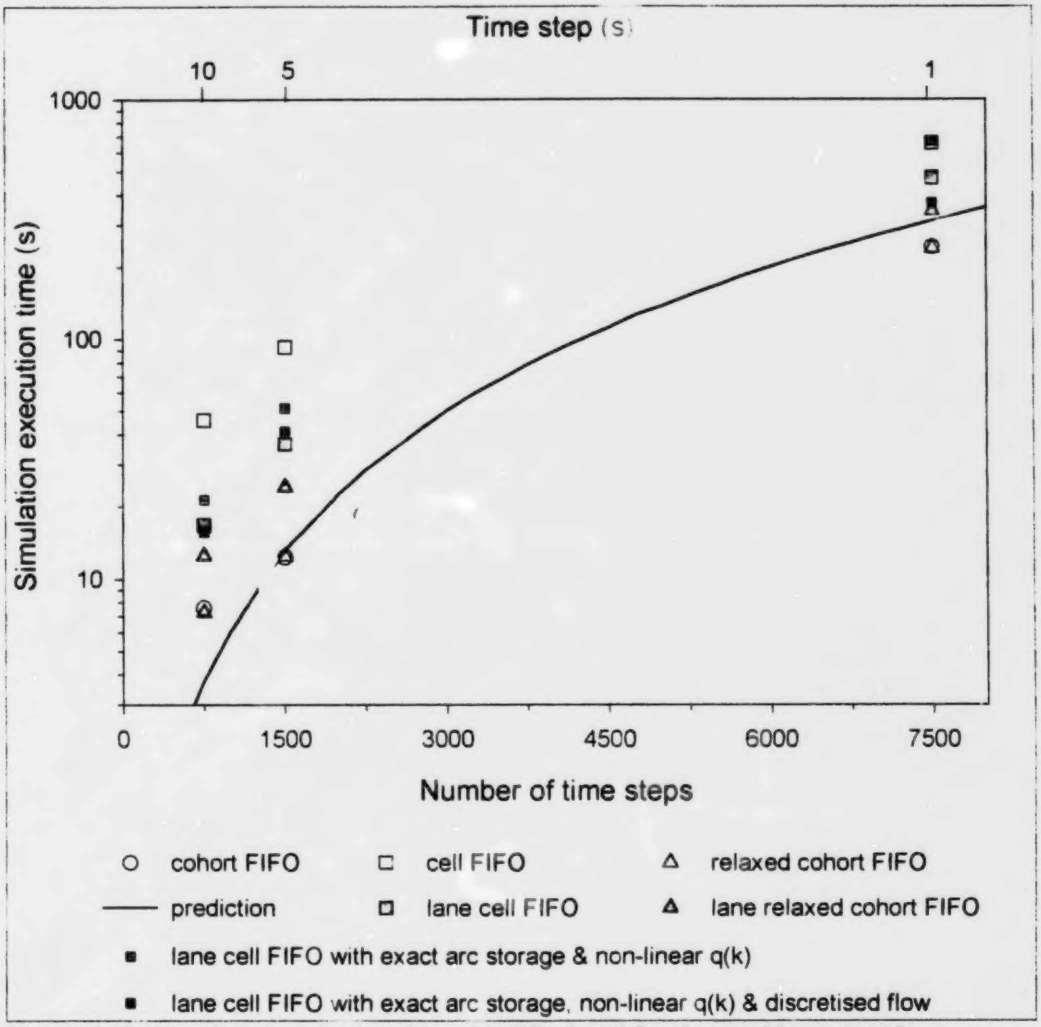


Figure 7.13 Simulation execution time of CellNetLoad for the corridor network

The algorithms with explicit lanes iterate to determine the partial flows for each cohort, until the local demand or the local supply is exhausted or when the lanes are blocked. More iterations are required to converge because the local demand and the local supply are distributed in proportion to the lane capacities, instead of the partial demands. The lane relaxed cohort FIFO algorithm (filled triangles) considers any spillback as part of the local demand, whereas the lane cell FIFO algorithm (filled squares) does not. Therefore, the cell FIFO algorithm takes longer to run.

The enhancements to the flow propagation model had less of an effect on the simulation execution time than the various intersection algorithms. The modelling of the exact arc storage reduced the amount of congestion in the network, hence decreasing the simulation execution time. The simulation with the non-linear equilibrium flow-density relationship required more time to evaluate the local supply and the local demand. The discretisation of flow has a bias which restrains flow, especially for small time steps and flows near capacity. This increase in the required simulation execution time is countered by the fact that no small packets are permitted on the network. The net effect for this network was a decrease in the simulation execution time.

7.5 Discussion

The intersection model must accurately determine the (partial) flows in order to accurately predict the temporal network flow, density, and travel time. The propagation of shockwaves depends entirely on the magnitude of the partial flows at an intersection. The improvements in accuracy may justify the significant time requirements for the simulation execution with complex intersection algorithms, depending on the application.

Based on comparisons with INTEGRATION, the strict enforcement of FIFO discipline, as in the LWR hydrodynamic model, is not realistic. This is expected since overtaking is permitted on any multi-lane roadway. Properly modelling overtaking is particularly important at the approach to an intersection, where lane striping restricts streams to certain lanes, allowing other streams to overtake in other lanes.

The effects of the intersection model are not restricted to the region closely surrounding the node. Shockwaves transmit the changes in flow and density at an intersection upstream and downstream. The shockwave speed can be properly modelled by the cell-transmission model only if the flow states on either side of the shockwave are known with some accuracy. The interaction of shockwaves, which may absorb or reflect one another, can alter and move congestion. Identifying the cause of a congested area may require carefully tracing back through the simulation results or forward from the known initial conditions.

The cell-transmission model cannot account for the capacity reduction caused by lane-changing manoeuvres. Though it is possible to predict the reduced capacity and then specify it in the input data. The prediction of the reduced capacity is difficult even in small networks with isolated sources of congestion. The propagation of congestion throughout a network with multiple causes of congestion makes predicting the temporal capacity reductions with any accuracy impossible. Of course, the temporal capacity reduction could be predicted by a microscopic simulation model with lane-changing logic, but that seems to render the cell-transmission model redundant.

Calibration and validation of a model must be carried out with all significant measures on a variety of networks. Although the dynamic network loading problem is primarily concerned with accurate predictions of the temporal path travel times, these are dependent on the arc travel times, which in turn are dependent on density, which is dependent on flow. An accurate temporal path travel time for a particular network, relative to some benchmark, is not sufficient. If the underlying flow and density do not match the benchmark, then any change to the network topology or the temporal demand may cause the estimated temporal path travel time to diverge from the benchmark.

Chapter 8

Conclusions and Recommendations

This chapter contains a summary of the theoretical contributions of this thesis. The conclusions found in the previous chapters of this thesis are synthesised. Comparisons are made between the findings and the objectives of this research. Subsequently, a list of recommendations and suggested future research is provided.

8.1 Contributions to theory

The first objective of this thesis is to produce a comprehensive explanation of the theoretical foundations of the cell-transmission model by integrating the contributions in the literature from diverse fields.

Chapter 2 contains a comprehensive explanation of the theory behind the cell-transmission model, which unites the contributions from the fields of fluid dynamics, thermodynamics, numerical methods, traffic flow theory, and traffic engineering. Whenever appropriate original insights were included. These insights are summarised below.

It was found that several modifications to the cell-transmission theory, which are used in practice, contradict assumptions in the theoretical foundation. The first such contradiction involves the definition of a shockwave and the use of an equilibrium flow-density relationship which has non-differentiable points. A concave equilibrium flow-density relationship also violates an assumption. Another is based on the discretisation of flow.

According to the entropy condition, which is equivalent to driver's ride impulse, a density discontinuity behaves either as a shockwave or as a fan. An increase in density across the discontinuity implies a shockwave, which remains as a discontinuity, assuming unbounded acceleration and deceleration rates. A decrease in density across the discontinuity implies a fan, which smoothes the discontinuity. On a space-time diagram, the characteristics fan out from a density (decrease) discontinuity. The fan can be visualised as the front of a dense platoon spreading into a downstream region of lower density.

The equilibrium flow-density relationship used in the cell-transmission model is often described as a piecewise linear function (Daganzo, 1994). If a density discontinuity spans a non-differentiable point of the piecewise linear function, such that the upstream density is larger than the downstream density, then no fan emerges. Instead, the fan is approximated as two spreading homogeneous flow states separated by a density discontinuity. This density discontinuity does not spread, so it behaves like a shockwave that violates the entropy condition. The approximation of the fan improves as the number of linear segments on the piecewise linear flow-density relationship increases.

The cell-transmission model is described as a discretised approximation to the LWR hydrodynamic model. The solution to the cell-transmission model converges to solution to the LWR hydrodynamic model as the discretisation is made finer. It has been suggested that the cell-transmission model could use a non-concave equilibrium flow-density relationship (Daganzo, 1999). Unfortunately, the proof for convergence of the solution of the cell-transmission model with the solution of the LWR hydrodynamic model assumes a concave flow-density relationship.

No proof is available which relaxes this assumption, so the use of a non-concave function should be approached with caution.

Several enhancements were made to CellNetLoad, an implementation of the cell-transmission model, in order to improve the realism of the simulation results relative to a benchmark. Several findings based on the implementation and analysis of the results are of a theoretical nature.

One fundamental assumption of the cell-transmission model is that flow is continuous. In reality, flow must have a discrete quantity, since vehicles are discrete. In addition, it is tempting to implement the cell-transmission model with a computer simulation that uses discrete vehicles. This has two advantages. Firstly, the limited precision for the representation of real numbers with a computer creates rounding errors that violate the conservation of flow. Using integer values for the quantity of vehicles resolves this issue simply. Secondly, the simulation may be slowed by the presence of many small quantities of vehicles on the network, which are eliminated if vehicles are represented discretely. It was found that the implementation of discretised flow in CellNetLoad results in a bias which restricts flow. This bias becomes more pronounced as the time discretisation is made finer and as the flows approach capacity. Therefore, the solution does not converge to the solution of the hydrodynamic model with finer discretisations.

Many traffic simulations model traffic signals statically. The capacity is reduced uniformly on a an approach according to the ratio of the duration of the green indication(s) to the cycle length. This results in two problems. Firstly, cyclic queuing does not occur, so density and travel time are underestimated, and spillback may be incorrectly modelled. A second problem is that streams which should operate independently, because they have separate green indications, may interact. These streams may impede one another as they merge on an arc in the forward star.

The traditional cell-transmission model permits only three types of cells: ordinary cells, diverge cells, and merge cells. Diverge cells must send flow to ordinary cells only. Merge cells must receive flow from ordinary cells only. These restrictions ensure that the exit flow from each cell depends only on the density of that cell and the density of the cells immediately downstream. Therefore, the calculation of flow may be performed in parallel for each cell, followed by the update of the density of each cell in parallel. The drawback of this restriction is that intersections with multiple arcs in the backward star and the forward star must be represented like an interchange with a diverge cell on each approach, which sends flow from each stream to a distinct ordinary cell. This representation requires the distribution of the approach capacity and storage among the ordinary cells. Spillback to the diverge cell affects the arrivals to the intersection, and interferes with shockwave propagation.

It was shown that the cell-transmission model can be implemented with a generalised intersection algorithm which permits diverge cells to send flow directly to merge cells. This relaxation respects all assumptions in the theory of the cell-transmission model. However, the potential parallelisation of the implementation is reduced since the flows through the intersection must be calculated simultaneously. The convergence of the generalised intersection algorithm was proven. Furthermore, the number of iterations required for termination is bounded from above if the discharge capacity and the local supply are distributed in proportion to the partial demands of the streams. For each set of partial demands, the number of iterations required for convergence of the algorithm is less than or equal to the number of streams with positive partial demand. It was also found that the generalised intersection algorithm is a flow maximising greedy algorithm, provided that the discharge capacity and the local supply are distributed in proportion to the partial demands of the streams.

The CellNetLoad implementation of the cell-transmission model has quadratic complexity. It depends linearly on the number of cells, arcs, time steps, and output intervals, which implies a dependence on the square of the inverse of the time step.

8.2 Findings and conclusions

The second objective of this thesis is to evaluate the performance of the cell-transmission model, and investigate enhancements which may improve the accuracy of the cell-transmission model. A qualitative evaluation is presented at the end of chapter 2. A quantitative comparison with a benchmark was used to evaluate the accuracy of the model. Several enhancements were developed and implemented, then compared with the original results and the benchmark.

An implementation of the cell-transmission model, called CellNetLoad, was validated on a series of test networks and the results were compared to a benchmark. The benchmark was derived from the simulation results of INTEGRATION, a microscopic traffic simulation program. Although the realism of INTEGRATION is certainly open to criticism, car following is explicitly modelled, as well as mandatory and discretionary lane changing. These phenomena are not modelled by the cell-transmission model. Instead, the cell-transmission model captures the effects of car following with a fluid dynamic model, and ignores lane changing altogether.

The car-following logic in INTEGRATION is based on a speed-headway relationship with bounded acceleration and deceleration. Shockwaves emerge from the car-following logic with speeds that respect the Rankine-Huginot jump condition of fluid dynamics. Due to the bounded acceleration and deceleration, shockwaves tend to dissipate, which is a realistic phenomenon.

The cell-transmission uses a flow-density relationship to propagate flow. The emergent shockwaves are consistent with the LWR hydrodynamic model as the time steps approach zero. Because acceleration and deceleration are unbounded, shockwave dissipation is due only to numerical error. A trapezoidal flow-density relationship tends to decrease the shockwave dissipation. However, the non-linear speed-headway relationship used in INTEGRATION has a corresponding non-linear flow-density relationship. Using this non-linear flow-density relationship in CellNetLoad increased the shockwave dissipation slightly, but reduced the discrepancy between the observed temporal flow, density, and travel time. No conclusions were made concerning the correctness of the functional form of the flow-density relationship, since it is based on curve fitting to traffic observation.

The discrepancy between the emergent shockwaves was overshadowed by other modelling discrepancies. The cell-transmission model ignores the impact of lane changing, which is a fundamental feature of traffic flow, because lanes and discrete vehicles are not represented. Vehicles are assumed to perform mandatory lane changes without penalty. In reality, lane changing is the mechanism which triggers congestion at lane reductions and when weaving and spillover occur. Lane changing results in a capacity reduction in terms of flow and storage, which depends on many factors, such as the flows, the speed, and aggressiveness of the drivers.

The lack of lane changing on the cell-transmission model impacts the modelling of lane drops, weaving, and spillover. The asymmetry of a lane drop from one side of the road is ignored. All lane-changes occur without penalty at the lane drop. The expected capacity reduction due to weaving is ignored. Spillover is modelled with FIFO discipline in the intersection algorithm. The effects of lane striping on spillover and the distribution of the discharge capacity and the local supply are ignored.

Some of these deficiencies can be compensated with enhancements to the cell-transmission model. A capacity reduction due to weaving and spillover can be specified in the input to CellNetLoad. A temporal capacity reduction, perhaps based on the current flow state, would be more effective at capturing the effects of weaving and spillover. Nevertheless, determining the magnitude of the capacity reduction accurately is not trivial. FIFO discipline may be relaxed to improve the modelling of spillover on multi-lane intersection approaches. Though the maximum relaxation of FIFO discipline permits overtaking among packets within the same cell only. As the time step is decreased, the cells become smaller, and the amount of overtaking is reduced. As the time step approaches zero, FIFO discipline is strictly enforced, which is consistent with the LWR hydrodynamic model. Thus, it can be concluded that the LWR hydrodynamic model is unrealistic for multi-lane intersection approaches. The modelling of spillover and the distribution of the discharge capacity and local supply at intersections can be improved further by explicitly modelling lane striping at intersection approaches.

The generalised intersection models improved the accuracy of the simulation of intersections. The relaxation on the configuration of diverge cells and merge cells allowed a more realistic representation of the intersection. In addition, the upper bound on the time step is increased up to 900 percent. The savings in execution time due to the permission of larger time steps is offset by the time-consuming nature of the algorithms relative to flow maximising algorithms for merges and diverges.

Discretising the flow is not recommended because of the bias which restrains the flow. The severity of the flow restraint is difficult to predict because it depends on the ratio of the flow to the capacity, on the size of the vehicle unit, and on the time step. Though some savings in the simulation execution were found due to the reduction in the number of small packets, the increased congestion due to the bias can drastically alter the simulation results and increase the simulation execution time.

8.3 Recommendations

In general, it was found that the cell-transmission was very good at propagating congestion on linear arcs accurately, provided that the flow states were correct on the upstream and downstream sides of the shockwave. The accuracy of the flow state depends heavily on the intersection model, which is particularly weak in the cell-transmission model. Enhancements aimed at improving the accuracy of the intersection algorithm met with limited success, because the cell-transmission model does not represent lanes or discrete vehicles, which are necessary for modelling lane changing.

It is possible to specify the predicted effects of lane changing as input to the cell-transmission model. However, accurately predicting these effects remains a challenge. Microscopic simulation programs currently offer the best estimates, but if these models are necessary to obtain the input of the cell-transmission model, then the cell-transmission model is redundant.

Parallelisation of the implementation and improvements to the simulation execution speed are important considerations for certain (real-time) applications, provided that the accuracy of the simulation results is of secondary importance.

References

Ansorge, R. What does the entropy solution mean in traffic flow theory? *Transportation Research*, Vol. 24B, pp. 133-43, 1990.

Astarita, V. A continuous time link model for dynamic network loading based on travel time function. *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. J.B. Lesort, Ed. Pergamon, Oxford, 1996.

Astarita, V. et al. Link based dynamic network loading models with spill-back: solution by simulation. Centre de Recherche sur les Transports, Université de Montréal, 1998a.

Astarita, V. et al. Link based dynamic network loading models with spill-back: intersection models. Centre de Recherche sur les Transports, Université de Montréal, 1998b.

Barceló, J., et al. Parallelization of microscopic traffic simulation for ATT sytems analysis. In *Equilibrium and Advanced Transportation Modelling*, P. Marcotte and S. Nguyen (Eds.), Kluwer Academic Publishers, 1998.

Ben-Akiva, M. and M. Bierlaire. DYNAMIT: A network-wide traffic prediction system. Transportation Research Board 78th Annual Meeting, Washington , D.C., 1999.

Bick, J.H. and G.F. Newell. A continuum model for two-directional traffic flow. *Quarterly Applied Mathematics*, Vol. 18, pp191-204, 1960.

Boyce, D.E., B. Ran, and L.J. LeBlanc. Solving an instantaneous dynamic user-optimal traffic assignment model. Urban Transportation Centre, University of Illinois at Chicago, ADVANCE Working Paper Series N.11, 1992.

Bui, D.D., P. Nelson, and S.L. Narasimhan. *Computational realizations of the entropy conditions in modelling congested traffic flow*. Report 1232-7. Texas Transportation Institute, 1992.

Carey, M. A constraint qualification for a dynamic traffic assignment model. *Transportation Science*, Vol. 20, pp. 55-8, 1986.

Carey, M. Optimal time-varying flows on congested networks. *Operations Research*, Vol. 35, 1987.

Carey, M. Nonconvexity of the dynamic assignment problem. *Transportation Research*, Pergamon Press, Vol. 26B, pp. 127-33, 1991.

Cayford, R., W.H. Lin, and C.F. Daganzo. The NETCELL simulation package: Technical description. California PATH Research Report UCB-ITS-PRR-97-23, University of California, Berkeley, 1997.

Chang G.L., H.S. Mahmassani, and R. Herman. A macroparticle simulation model to investigate peak-period commuter decision dynamics. *Transportation Research Record*, Vol. 1005, pp 107-20, 1985.

Codina, E. and J. Barceló. Dynamic traffic assignment: Considerations pn some deterministic modelling approaches. *Annals of Operations Research*, Vol. 60, pp. 1-58, 1995.

Cohen, S. Kinematic wave theory. In *Concise Encyclopedia of Traffic & Transportation Systems*. M. Papageorgiou, Ed. Pergamon Press, pp. 231-234, 1991.

Courant, R., K.O. Friedrichs, and H. Lewy. Uber die partiellen differenzengleichungen der physik. *Math. Ann*, Vol. 100, p. 32, 1928.

Cremer, M. and A.D. May. An extended traffic model for freeway control. *Research Report UCB-ITS-RR-85-7* Berkeley, California: Institute of Transportation Studies, University of California, 1985.

Daganzo, C.F. The cell-transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research*, Pergamon Press, Vol. 28B, pp. 269-87, 1994.

Daganzo, C.F. The cell-transmission model, part II: Network traffic. *Transportation Research*, Pergamon Press, Vol. 29B, pp. 79-93, 1995a.

Daganzo, C.F. A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research*, Pergamon Press, Vol. 29B, pp. 261-276, 1995b.

Daganzo, C.F. A continuum theory of traffic dynamics for freeways with special lanes. *Transportation Research*, Pergamon Press, Vol. 31B, pp. 83-102, 1997a.

Daganzo, C.F. A simple physical principle for the simulation of freeways with special lanes and priority vehicles. *Transportation Research*, Pergamon Press, Vol. 31B, pp. 103-125, 1997b.

Daganzo, C.F. The lagged cell-transmission model. In *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*. A. Ceder, Ed. Pergamon Press, pp 81-104, 1999.

Drake, J.S., J.L. Schofer, and A.D. May, Jr. A statistical analysis of speed density hypotheses. *Proceedings of the 3rd International Symposium on Transportation and Traffic Theory*, Elsevier North Holland, Inc., New York, 1967.

Drissi-Kaïtouni, O. A variational inequality formulation of the dynamic traffic assignment. Centre de Recherche sur les Transports, Université de Montréal, Publication No. 735, 1990.

Drissi-Kaitouni, O. and A. Hameda-Benchekroun. A dynamic traffic assignment model and a solution algorithm. *Transportation Science*, Vol. 26, 1992.

Edie, L.C. Following and steady-state theory for non-congested traffic. *Operations Research*, Vol. 9, pp. 66-76, 1961.

Fellendorf, M. VISSIM: A microscopic simulation tool to evaluate actuated signal control including bus priority. Presented at the *64th ITE Annual Meeting*, session 32, Dallas, Oct. 1994.

Ferrer, J.L. and J. Barceló. AIMSUN2: Advanced interactive microscopic simulator for urban and non-urban networks. Internal report. Departamento de Estadistica e Investigación Operativa, Facultad de Informática, Universitat Politècnica de Catalunya, 1993

Florian, M and D. Hearn. Network equilibrium models and algorithms. In *Handbooks in OR & MS: Network Routing*. M.O. Ball et al. (Eds.). Elsevier, Vol. 8, pp. 485-550, 1995.

Friesz, T.L. , F.J. Luque, R.L. Tobin, and B.W. Wie. Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research*, Vol. 37, pp. 58-69, 1989.

Gabard, J.F. Car-following models. In *Concise Encyclopedia of Traffic and Transportation Systems*, M. Papageorgiou (Ed.), Pergamon Press, Oxford, 1991.

Gazis, D.C. *Traffic Science*. John Wiley, 1974.

Godunov, S. A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics. *Mat Sb*. Vol 47, pp. 271-306, 1959.

Greenberg, H. An analysis of traffic flow. *Operations Research*, Vol. 7, pp. 78-85, 1959.

Greenshields, B.D. A study in highway capacity. *Proceedings of the Highway Research Board.* Vol 1, p. 458, 1935.

Helbing, D. *Verkehsdynamik.* Springer, 1997.

Jayakrishnan, R., H.S. Mahmassani, and Ta-Yin Hu. An evaluation tool for advanced traffic information and management systems in urban networks. *Transportation Research*, Vol. 2C, pp. 129-47, 1994.

Kerner, B.S., P. Konhaüser, and M. Shike. A new approach to problems of traffic flow theory. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory.* J.B. Lesort, Ed. Pergamon, Oxford, pp. 79-102, 1996.

Khattak, A.J. and P.P. Jovanis. Capacity and delay estimation for priority unsignalised intersections: Conceptual and empirical issues. *Transportation Research Record*, Vol. 1287, Transportation Research Board, National Research Council, Washington, D.C., pp. 129-37, 1990.

Lax, P.D. Weak solutions of nonlinear hyperbolic equations and their numerical computations. *Communications on Pure and Applied Mathematics*, Vol. 7, pp. 159-93, 1954.

Lax, P.D. Shock waves and entropy. *Contributions to Nonlinear Functional Analysis.* E. Zarantonello, Ed. New York, Academic Press, 1971.

Lax, P.D. Hyperbolic systems of conservation laws and the mathematical theory of shock waves. *Regional Conference Series in Applied Mathematics*, Vol. 11. Philadelphia, PA, Siam, 1973.

Lebacque, J.P. The Godunov scheme and what it means for first order traffic flow models. *Transportation and Traffic Theory: Proceedings of the 13th International Symposium.* J.B. Lesort, Ed. Pergamon, pp. 645-77, 1996.

Lebacque, J.P and J.B. Lesort. Macroscopic traffic flow models: A question of order. *Transportation and Traffic Theory: Proceedings of the 14th International Symposium.* A. Ceder, Ed. Pergamon, pp. 3-25, 1999.

Li, Y.I., A. Ziliaskopoulos, and S.T. Waller. Linear programming formulations for system optimum dynamic traffic assignment with arrival time based and departure time based demands. Presented at *Transportation Research Board 78th Annual Meeting*, 1999.

Lighthill, M.J. and G.B. Whitham. On kinematic waves II: A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London, series A.* Royal Society, Vol. 229, pp. 317-345, 1955.

Liu, R. and D. van Vliet. DRACULA – a dynamic microscopic model of road traffic. *Proceedings of the International Transport Symposium*, pp. 160-170, Beijing, July 1996.

Lo., H.K. A dynamic traffic assignment formulation that encapsulates the cell transmission model. *Transportation and Traffic Theory: Proceedings of the 14th International Symposium.* A. Ceder, Ed. Pergamon, pp. 327-350, 1999.

Luke, J.C. Mathematical models for landform evolution. *Journal of Geophysical Research*, Vol. 77, pp. 2460-4, 1972.

Mahut, M. Thesis proposal. Departement d'Informatique et de Recherche Opérationnelle, , Université de Montréal, 1999.

May, A.D. *Traffic Flow Fundamentals.* Prentice Hall, 1990.

Mauro, V. Evaluation of dynamic network control: simulation results using NEMIS urban simulator. Ttansportation Research Board 70th Annual Meeting. Washington, D.C., 1991.

Merchant, D.K. and G.L. Nemhauser. A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science*, Vol. 12, 1978a

Merchant, D.K. and G.L. Nemhauser. Optimal conditions for dynamic traffic assignment model. *Transportation Science*, Vol. 12, 1978b.

Michalopoulos, P.G., D.E. Beskos, and J.K. Lin. Analysis of interrupted traffic flow by finite difference methods. *Transportation Research*, Vol. 18B, pp. 409-21, 1984a.

Michalopoulos, P.G., D.E. Beskos, and Y. Yamauchi. Multilane traffic flow dynamics: Some macroscopic considerations. *Transportation Research*, Vol. 18B, pp. 377-96, 1984b.

Munjal, P.K. and L.A. Pipes. Propagation of on-ramp density perturbations on unidirectional two- and three-lane freeways. *Transportation Research*, Vol. 5, pp. 241-55, 1971.

Nagel, Kai. From particle hopping models to traffic flow theory. *Transportation Research Board 77th Annual Meeting*. Preprint No. 981331. 1998.

Nagel, Kai, et al.. TRANSIMS traffic flow characteristics. *Transportation Research Board 77th Annual Meeting*. Preprint No. 981332. 1998.

Newell, G.F. A simplified theory of kinematic waves in highway traffic. *Transportation Research*, Vol. 27B, pp. 281-314, 1993.

Omli, R. and J.L. Farges. DMRG simulation: Specification for microscopic models. PROMETHEUS DMRG Workshop. Toulouse, July 1992.

Osher, S. Riemann solvers, the entropy condition, and difference approximations. *SIAM Journal of Numerical Analysis*, Vol. 21, pp. 217-235, 1984.

Payne, H.J. Models of freeway traffic and control. In *Mathematical Models of Public Systems, Simulation Council Proceedings*. Vol 28, pp. 51-61, 1971.

Payne, H.J. FREFLO. A macroscopic simulation model of freeway traffic. *Transportation Research Record*, Vol 722, pp. 68-77, 1979.

Phillips, W.F. A kinetic model of traffic flow with continuum implications. *Transportation Planning and Technology*. Vol. 5-3, pp. 131-8, 1979.

Prigogine, I. and R Hermann. *Kinetic Theory of Vehicular Traffic*. American Elsevier, New York, 1971.

Ran, B., D.E. Boyce, and L.J. LeBlanc. Dynamic user-optimal traffic assignment model: A new model and solution technique. Presented at the *1st Triennial Symposium on Transportation Analysis*, Montréal, Canada, 1991.

Richards, P.I. Shock waves on the highway. *Operations Research* Operations Research Society of America, Vol. 4, pp. 42-51, 1956.

Richtmyer, R.D. and K.W. Morton. *Difference methods for initial value problems*. Interscience, New York, 1967.

Sawaya, O.B., A. Ziliaskopoulos and J.L. Schofer. Development of dynamic control strategies for managing freeway incident traffic. Presented at *Transportation Research Board 78th Annual Meeting*, 1999.

Special Report 209: Highway Capacity Manual, 3rd ed.. Transportation Research Board, National Research Council, 1998.

Underwood, R.T. Speed, volume, and density relationships, quality and theory of traffic flow. Yale Bureau of Highwau Traffic, New Haven, Connecticut, pp. 141-88, 1961.

Van Aerde, M. A single regime speed-flow-density relationship for freeways and arterials. Presented at the Transportation Research Board 74th Annual Meeting, Washington, D.C., 1995.

Van Aerde, M. INTEGRATION Release 2.20 for WINDOWS: User's Guide. May 1999a.

Van Aerde, M. Personal communication. May 1999b.

Van Aerde, M. Single regime speed-flow-density relationship for congested and uncongested highways. Transportation Research Board 74th Annual Meeting, Washington , D.C., 1995.

Velan, S.M., and M. Van Aerde. The impact of driver and flow variability on capacity estimates of permissive movements. *Transportation Research*, Vol. 32A, 1998.

Vvedenskaya, N.D. Solution of the Cauchy problem for a non-linear equation with discontinuous initial conditions by the method of finite differences. *Dokl. Akad. Nauk SSSR (N.S.)*, Vol. 111, pp. 517-20, 1956.

Wie, B.W., T.L. Friesz, and R.L. Tobin. Dynamic user optimal traffic assignment on congested multidestinational networks. *Transportation Research*, Vol. 24B, 1990.

Xu, Y.W., J.H. Wu, M. Florian, P. Marcotte, and D.L. Zhu. Advances in the continuous dynamic network loading problem. *Transportation Science*, Vol. 33, pp. 341-53, 1999.

Yang, Q. and H. Koutsopoulos. A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research*, Vol. 4C, pp. 113-29, 1996.

Yedlin, M., E. Lieberman, A. Phlegar, A. Kanaan, and A. Santiago. The new TRAF-NETSIM: Version 5.0. Presented at the Transportation Research Board 73rd Annual Meeting, Washington, D.C., 1994.

Ziliaskopoulos, A. and S. Lee. A cell transmission based assignment-simulation model for integrated freeway/surface street systems. *Transportation Research Record*, 1997.

Appendix.

Format of Input Files for INTEGRATION

Table A1 Description of the fields in the master control file – Format III

Line	Field	Description
1	1	File title (up to 40 characters) The title will appear in formatted output and can serve to clearly identify the results.
2	1	Total simulation time (seconds) [real] $(x \geq 1)$
	2	Frequency of outputs (seconds) to output file 10 [integer] $(0 \leq \lvert x \rvert \leq$ *simulation time*$)$ If $x = 0$ then no output is written to file 10. If $x > 0$ then output is written to file 10 at time zero and every x seconds. If $x < 0$ then no output is written to file 10 at time zero, but such output is written at every $\lvert x \rvert$ seconds there after.
	3	Frequency of outputs (seconds) to output files 12 - 14 [integer] $(0 \leq \lvert x \rvert \leq$ *simulation time*$)$ If $x = 0$ then no output is written to files 12-14. If $x > 0$ then output is written to the files at time zero and every x seconds. If $x < 0$ then no output is written to the files at time zero, but such output is written at every $\lvert x \rvert$ seconds there after.
	4	Master File Routing Format [integer] For routing format II, $x = 2$
	5	Simulation termination [integer] If $x = 0$ then animation screen closes at end of regular simulation. If $x = 1$ then animation screen pauses just prior to end of simulation.
3	1-5	Minimum Path Construction Method [integer] If $x = 1$, Macroscopic User Equilibrium. If $x = 2$, Microscopic Feedback for Sub-populations If $x = 3$, Microscopic Feedback for Individuals, and If $x = 4$, Microscopic Dynamic Traffic Assignment
4	1-5	Frequency (seconds) at which driver classes 1 through 5 respectively, will re-compute their minimum paths pre-trip [integer] $(0 \leq x \leq$ *simulation time*$)$ If $x = 0$ then minimum paths are only computed at time zero. If $x > 0$ then minimum paths are re-computed on average every x seconds
5	1-5	Frequency (seconds) at which driver classes 1 through 5 respectively, will re-compute their minimum paths en-route [integer] $(0 \leq x \leq$ *simulation time*$)$ If $x = 0$ then minimum paths are only computed at time zero. If $x > 0$ then minimum paths are re-computed on average every x seconds.
6	1-5	Information source for driver classes 1 through 5 respectively [integer] If $x = 1$ then paths are updated based on real-time information If $x = 2$ then paths are updated based on data from file 8. If $x = 3$, then paths are updated based on data from file 9.
7	1-5	Update weight for real-time data base for vehicles of same type [integer] $(0 \leq x \leq 100)$
8	1-5	Update weight for real-time data base for vehicles of same type [integer] $(0 \leq x \leq 100)$
9	1-5	Travel time assumption for vehicles not under surveillance. [integer] If $x = 1$ then travel times are assumed to be at free-speed If $x = 2$ then travel times are assumed to be as listed in file 6. If $x = 3$, then travel times are assumed to be as listed in file 7.
10	1-5	Amount of error introduced into the link travel time data, prior to the calculation of the minimum path trees [real] $(0.0 \leq \lvert x \rvert \leq 0.5)$, where x equals COV (standard deviation/mean) of link travel time. If $x = 0$ then no error is introduced. If $x \geq 0$ then error is considered to be normally distributed. If $x \leq 0$ then error is considered to be log-normally distributed.
11	1	Directory path for directory in which the input files are located. If the line is left blank, then the model assumes that the files are in the current directory.

12	1	Directory path for directory in which output files are to be placed. If the line is left blank, then the model assumes that the files are in the current directory. Note that path must be different from the input subdirectory.
13-17	1	Names of required input data files (Files 1 - 5)
18-23	1	The name `none` or `NONE` should appear in these fields.
24	1	Name of the required standard output file (File 10) The name `none` or `NONE` should appear if the optional file is not utilized, however, a default file called `default.u10` will be created.
25-35	1	Names of the optional input/output files (Files 11-22) The name `none` or `NONE` should appear if the optional output file is not desired.

Table A2 Description of fields in node file

Line	Field	Description
1	1	File title (up to 40 characters)
2	1	Number of nodes for which information is provided in the file [integer] $(2 \le x \le \textit{max. number of nodes})$
	2	Scale factor applied to all x-coordinate values listed in the file [real] $(x > 0.0)$ This factor can be used to convert coordinates from any unit system into the desired units of kilometers.
	3	Scale factor applied to all y-coordinate values listed in the file [real] $(x > 0.0)$ This factor can be used to convert coordinates from any unit system into the desired units of kilometers.
3+	1	Unique node identification number [integer] $(1 \le x \le \textit{max. node number})$ Node numbers cannot be duplicated and should be consistent with those in files 2 and 4
	2	X-coordinate of node location [real] $(x > 0.0)$
	3	Y-coordinate of node location [real] $(x > 0.0)$
	4	Node type identifier [integer] $(1 \le x \le 4)$ 1 = both origin and destination zone 2 = destination zone only 3 = origin zone only 4 = intermediate node only
	5	Macro-zone cluster number [integer] $(\lvert x \rvert \le \textit{max. macro-zone number})$ $x = 0$ for node type 3 or 4. $x < 0$ when zone is the centroid of macro-zone cluster x (node type 1 or 2 only).
	6	Information availability indicator [real] Recommend $x = 0$ then no information providing device exists at that node. (See *Volume 2* for more details)
	7	Optional alpha-numeric descriptor of this node (maximum 40 characters)

Table A3 Description of fields in link characteristics file

Line	Field	Description
1	1	File title (maximum 40 characters)
2	1	Number of links listed in the file [integer] ($2 \le x \le$ *max. number of links*)
	2-6	Scale factors for link length, free-speed, saturation flow rate, speed at capacity, and jam density, respectively [real] ($x > 0.0$)
3+	1	Unique link identification number [integer] ($1 \le x \le$ *max. link number*)
	2	Identification number of the node at the upstream end of the link [integer] ($1 \le x \le$ *max. node number*) The node number must be contained within file 1.
	3	Identification number of the node at the downstream end of the link [integer] ($1 \le x \le$ *max. node number*) The node number must be contained within file 1
	4	Link length (km) [real] (*min. length* $\le x \le 6.000$), where *min. length* = 0.010×number of lanes on the link
	5	Free speed on link (km/h) [real] ($10.0 \le x \le 200.0$)
	6	Basic saturation flow rate per lane (vph) [real] ($100.0 \le x \le 10000.0$)
	7	Number of lanes [real] ($1.0 \le x \le 7$)
	8	Vehicle speed variability factor [real] Recommend $x = 0$. (See *Volume 2* for more details).
	9	Speed at capacity (km/h) [real] ($10.0 \le x \le$ *free speed*)
	10	Jam density (veh/km/lane) [real] (*density at capacity* $\le x \le 5$**density at capacity*), where *density at capacity* = saturation flow/speed at capacity
	11	Turn prohibition indicator [integer] ($0 \le x \le$ *max. link number*) If $x = 0$ then no turns are prohibited. If $x > 0$ then x is interpreted as the identification number of the link to which turns from the present link are not permitted.
	12	Simulation time at which the prohibition of field 11 is activated [integer]
	13	Simulation time at which the prohibition of field 11 is deactivated [integer]
	14	Link number of first link opposing current link [integer] ($0 \le x \le$ *max. link number*) If $x = 0$ then no links oppose the current link. If $x > 0$ then x is interpreted as the number of the first link which opposes the flow of the current link.
	15	Link number of second link opposing current link [integer] ($0 \le x \le$ *max. link number*) If $x = 0$ then no second link opposes the current link. If $x > 0$ then x is interpreted as the number of the second link which opposes the flow of the current link.
	16	Number of the traffic signal which controls exit privileges of this link [integer] ($x \le$ *max. signal number* or $x = 10001$ or 10002) If $x = 0$ then no signal controls exit privileges. If $x > 0$ then x is the traffic signal number controlling exit from this link. If $x = 10001$ then a stop sign is modeled. If $x = 10002$ than a yield sign is modeled.
	17	Number of the first phase of the above signal in which this link discharges [integer] ($1 \le x \le$ *max. number of phases*) If $x = 0$ then signal does not exist.

	18	Number of the second phase in which this link discharges [integer] (0 ≤ *x* ≤ *max. number of phases*) If *x* = 0 then signal does not exist or link does not discharge during a second phase.
	19	Vehicle type link prohibition indicator [integer] (00000 ≤ *x* < 11111) Recommend *x* = 00000 indicating no prohibitions exist
	20	Surveillance level indicator [integer] (00000 ≤ *x* < 11111) Recommend *x* = 11111 to indicate that link is under surveillance
	21	Optional descriptive name of link [alpha-numeric]

Table A4 Description of fields in the signal timing file

Line	Field	Description
1	1	File title (maximum 40 characters)
2	1	Number of traffic signals within the traffic network [integer] (0 ≤ *x* ≤ *max. number of signals permitted*)
	2	Number of traffic signal plans provided in this file for each signal [integer] (0 ≤ *x* ≤ 1000)
	3	Duration of each signal timing plan (seconds) [integer] (*x* ≥ 60, *where x is a multiple of 60*)
3	1	Signal plan number [integer] (1 ≤ *x* ≤ *number of traffic signal plans specified*) Signal plans must be numbered sequentially in increasing order.
4+	1	The traffic signal number for which the remaining fields on this line apply [integer] (1 ≤ *x* ≤ *max. signal number*)
	2	Initial cycle length (seconds) that will be utilized in the simulation [real]
	3	Minimum cycle length (seconds) that will be permitted in the event that signal timing optimization is performed [real] (0.0 ≤ *x* ≤ *max. cycle length*)
	4	Maximum cycle length (seconds) that will be permitted in the event that signal timing optimization is performed [real] (*x* ≥ *min. cycle length*)
	5	Time in seconds, after the start of the current time plan, that the first phase of this signal begins. This time in normally referred to as the signal offset [integer] (0 ≤ *x* ≤ *cycle length*)
	6	Number of phases at this traffic signal [integer] (2 ≤ *x* ≤ 8)
	For each of the phases specified in field 6 the following two fields must be provided	
		Effective green time (seconds) of this phase [real] (1.0 ≤ *x* < *cycle time specified*)
		Effective lost time (seconds) of this phase [real] (0.0 ≤ *x* ≤ 10.0)
	last	Frequency (seconds) at which the phase splits and cycle lengths of this signal will be optimized [integer] (*x* ≥ 0, *where x is a multiple of 60*)

Table A5 Description of the fields in the demand file

Line	Field	Description
1	1	File title (maximum 40 characters)
2	1	Number of O-D demands to be read [integer] ($0 \leq x \leq$ *max. number of O-D pairs*)
	2	The number of the first O-D pair loading to be utilized by the model [integer] ($0 \leq x \leq$ *max. O-D number specified*) If $x = 0$ then x is set to be 1.
	3	The number of the last O-D pair loading to be utilized by the model [integer] ($0 \leq x \leq$ *max. O-D number specified*) If $x = 0$ then x is set to be *max. O-D number specified*.
	4	Global O-D demand scaling factor [real] ($x > 0.0$) For example, if x equals 2.0 then all listed departure rates are assumed to be double their listed values.
3+	1	O-D line identification number [integer] ($1 \leq x \leq$ *number of O-D pairs specified*)
	2	Origin zone identification number [integer] ($1 \leq x \leq$ *max. zone number*)
	3	Destination zone identification number [integer] ($1 \leq x \leq$ *max. zone number*)
	4	Departure rate (vph) for given O-D [real]
	5	The fraction of the vehicle headway that is random [real] ($0.0 \leq x \leq 1.0$)
	6	Time (seconds) at which specified O-D demand begins [integer] ($x \geq 0$)
	7	Time (seconds) at which specified O-D demand ends [integer] ($0 \leq x \geq$ start of O-D demand)
	8-12	The fraction of the demand that is to be of driver class i [real] ($0.0 \leq x_i \leq 1.0$) $\sum_{i=1}^{5} x_i = 1.0$
	13	The fraction of the total demand that will act as vehicle probes for the purposes of generating output for optional output files 15 and/or 16 [real] ($x \leq 1.0$) Recommend an initial value of $x = 0.0$.
	14	The capacity-based passenger car equivalency for all vehicles of this O-D demand [real] ($0.5 \leq x \leq 10.0$)

Table A6 Description of the fields in the incident file

Line	Field	Description
1	1	File title (maximum 4[illegible] characters)
2	1	Number of incident descriptor records to be read [integer] ($0 \le x \le$ *max. number of incidents permitted*)
3+	1	Incident record number [integer] ($1 \le x \le$ *max. incident number*)
	2	Identification number of the link in which the incident will occur [integer] ($1 \le x \le$ *max. link number*)
	3	Distance along the link (km), measured from the upstream end, where the incident starts [real] ($0 < x \le$ *length of link*)
	4	Distance along the link (km), measured from the upstream end, where the incident ends [real] (*incident start point* $< x \le$ *length of link*)
	5	Simulation time (seconds) at which incident is to begin [integer]
	6	Simulation time (seconds) at which the incident is to end [integer]
	7	Lane blockage indicator [binary] (x [illegible]) An n-digit number indicating the status of each of the n lanes on the link (0 accessible or 1 blocked), coding the lanes from median to shoulder.

Table A7 Description of fields in the optional lane striping input file

Line	Field	Description
1	1	File title
2	1	Number of records defined in the file [integer] ($0 < x \le$ *max. number of links in network*)
3	1	Record number identification [integer] ($0 < x \le$ *max. number of records in file*)
	2	Link number on which permission/prohibition exists [integer] ($0 < x \le$ *max. link number in network*)
	3	Number of lanes n on current link, as stated in file 2 [integer] ($0 < n \le$ *maximum number of permissible lanes*) Note, lanes are numbered consecutively from 1 to n from the median lane to the shoulder lane.
	4+	For each lane n, a three digit number xyz indicating permitted movements (0 prohibited or 1 permissible) [integer], where: x: left turn permission [binary] y: straight through permission [binary] z: right turn permission [binary] Note, there are 7 (2^3-1) acceptable values of xyz (000 not acceptable)
	4+n	For each lane n, a five digit number *abcde* indicating lane use prohibition by vehicle type (0 permitted or 1 prohibited) [integer] where: a: vehicle type 1 prohibition [binary] b: vehicle type 2 prohibition [binary] c: vehicle type 3 prohibition [binary] d: vehicle type 4 prohibition [binary] e: vehicle type 5 prohibition [binary] Note, there are 31 (2^5-1) acceptable values of *abcde* (11111 not acceptable)

VITA

Name: Shane Michael Velan

Place and year of birth: Montreal, 1973

Education:

Université de Montréal, 1997-2000
Ph.D. candidate (Computer Science)

Queen's University, 1995-1997
M.Sc. (Civil Engineering)

Queen's University, 1991-95
B.Sc. (Civil Engineering) 1995

Experience:

Teacher's Assistant
Dept. of Civil Engineering, Queen's University, 1995-1997

Traffic Analyst, Summer Intern
Kittelson & Associates, Inc., Portland, Oregon, Summer 1996

Research Assistant
Dept. of Civil Engineering, Queen's University, Winter 1996

Research Assistant, NSERC Research Award
Natural Sciences and Engineering Research Council
Dept. of Civil Engineering, Queen's University, Summer 1994

Awards:

NSERC Postgraduate Scholarship (PGS B), 1997-99

NSERC Postgraduate Scholarship (PGS A), 1995-97

John Vardon Memorial Scholarship,
Institute of Transportation Engineers, 1996

Queen's Graduate Award, Queen's University, 1996

Student Paper Award, Institute of Transportation Engineers, 1995

Undergraduate Student Paper Competition - First Prize,
Canadian Society for Civil Engineering - Transportation Division, 1995

The Zurbrigg Memorial Scholarship, Queen's University, 1995

S.D. Lash Scholarship, Queen's University, 1995

Provincial Scholarship, Queen's University, 1991-95

Canada Scholarship, Queen's University, 1991-95

Dean's Scholar, Queen's University, 1991-95

Senator Frank Carrel Upper-Year Scholarships,
Queen's University, 1991-94

Publications:

S. Velan. *Gap Acceptance of Permissive Movements at Signalised and Unsignalised Intersections*. Master's thesis. Department of Civil Engineering, Queen's University, 1997.

S. Velan and M. Van Aerde. The impact of driver and flow variability on capacity estimates of permissive movements. *Transportation Research*, Vol. 32A, 1998.

S. Velan and M. Van Aerde. Gap Acceptance and Approach Capacity at Unsignalized Intersections. ITE Journal, Vol. 66, pp. 40-5, March 1996.

S. Velan and M. Van Aerde. Relative Effects of Opposing Flow and Gap Acceptance Characteristics on Approach Capacity at Unsignalized Intersections. Transportation Research Board Annual Conference, 1996.